October 2019

Effects of Phonological Contrast on Within-Category Phonetic Variation

Ivy Hauser

EFFECTS OF PHONOLOGICAL CONTRAST ON WITHIN-CATEGORY PHONETIC VARIATION

A Dissertation Presented

by

IVY HAUSER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2019

Linguistics

EFFECTS OF PHONOLOGICAL CONTRAST ON WITHIN-CATEGORY PHONETIC VARIATION

A Dissertation Presented

by

IVY HAUSER

Approved as to style and content by:

Kristine Yu, Co-chair

John Kingston, Co-chair

Joe Pater, Member

Meghan Armstrong, Member

Joe Pater, Department Chair
Linguistics

DEDICATION

To my parents, Tammy and David Hauser.

ACKNOWLEDGMENTS

I am immensely grateful for the mentoring I have received from my dissertation committee, especially the co-chairs, Kristine Yu and John Kingston. Kristine and John have provided many years of advising, starting with serving as specialists on my first generals paper committee. I am thankful that we have continued to work together, as their advising styles and areas of expertise have caused me to think more critically about linguistics and be more confident as a scholar. I'm thankful for the amount of time and energy Kristine and John have put into mentoring me over the years: supporting me in research, teaching, professional, and personal development.

I would also like to thank the other members of my committee, Joe Pater and Meghan Armstrong. I started my career at UMass working fairly closely with Joe. As my interests starting heading more towards phonetics, Joe continued to provide advising on my research and professional development. I am especially grateful for his big-picture perspective at times when I have been stuck in the details of experimentation. Thanks to Meghan Armstrong for feedback and advice throughout my time at UMass. I appreciate her serving as a member on this committee and for her outside perspective on the research.

This work could not have been done without the help of several undergraduate research assistants who have put in many hours of textgridding, recording participants, gathering word frequency data, browsing dictionaries, hand-checking alignment, and listening to audio files. I would like to thank Greg Feliu for his work with Hindi and English experiment design and analysis, Saumya Joshi for her work with Hindi participants, Diana Konarski for her work with Polish experiment design and participants, Noah Constant for his work with French experiment design and participants,

Zachary Sun for his work with Mandarin experiment design and participants, and Allison Chen for her work with Mandarin participants.

There are several other faculty members at UMass who have provided important feedback and support throughout the dissertation process. I'd like to thank Gaja Jarsoz for her input on this work, asking questions that have made me think about my hypotheses in a completely different way. I also appreciate her assistance with my experiment on Polish sibilants (Chapter 3). Aside from the dissertation, Gaja has been an excellent mentor with regards to my research, my teaching, and my professional development. I'd also like to thank John McCarthy who was one of my primary advisors during my first few years at UMass.

Thanks to the staff in the UMass Linguistics department, Tom Maxfield and Michelle McBride from providing administrative support around the dissertation (and graduate school in general). They have been understanding and patient with my many questions from me about scheduling, paperwork, reimbursement, and more. Thanks for helping the non-research parts run smoothly.

Funding for this project came from the National Science Foundation, the University of Massachusetts Graduate School, and the University of Massachusetts Linguistics Department. I would like to thank Joe Pater, John McCarthy, Kristine Yu, John Kingston, Kyle Johnson, Heidi Bauer-Clapp, Tom Maxfield, and the University of Massachusetts Office of Grant and Contract Administration for their assistance with the NSF Graduate Research Fellowship and NSF Doctoral Dissertation Research Improvement Grant. The author is supported by the National Science Foundation Graduate Research Fellowship under Grants No. 1451512 and 823869. This material is based upon work supported by the National Science Foundation under Grants No. 1451512 and 823869. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Many thanks to Jean-Luc Schwartz, Louis-Jean Boë, Pierre Badin, and Thomas R. Sawallis for sharing the data from their 2012 paper and providing both scripts and helpful commentary. These data are used for the analysis in Chapter 5.

I would like to thank my cohort here at UMass: Caroline Andrews, Sakshi Bhatia, David Erschler, Coral Hughto, Jyoti Iyer, Leland Kusmer, and Katerina Vostrikova. Thanks for support, friendship, and feedback. Special thanks to Sakshi Bhatia and Jyoti Iyer for being wonderful housemates and Coral Hughto for being an excellent office mate. There are many other UMass students and alumni who have supported me during this process. Thanks to those who have heard me talk about my work countless times and offered feedback at PRGs and Sound Workshops: Brandon Prickett, Katie Tetzloff, Max Nelson, Andrew Lamont, Amanda Rysling, Aleksei Nazarov, Claire-Moore Cantwell, Brian Smith, Presley Pizzo, and Robert Staubs. Thanks to all the students in the UMass community who have provided friendship and made for a great working environment.

The Office of Professional Development at the University of Massachusetts provided support at various stages of the dissertation process. Thanks especially to Johanna Yunker and Heidi Bauer-Clapp whose workshops and advice have improved my writing, teaching, grant applications, job applications, time management, and confidence. Thanks also to The National Center for Faculty Development and Diversity. Their seminars and materials have helped me optimize productivity with this dissertation and have supported my general well-being as a scholar.

Thanks to audiences at the University of Massachusetts Linguistics Department who have provided feedback when I presented parts of this dissertation in Sound Workshop, Psycholinguistics Workshop, Joint Lab Meeting, the Phonology Reading Group, and various courses. Thanks also to audiences at the Linguistic Society of America 2016 meeting, the June 2017 meeting of the Acoustical Society of America, the May 2018 meeting of the Acoustical Society of America, the 2018 Annual Meeting

on Phonology, the Workshop on Phonological Variation and its Interfaces, the 2019 meeting of the American Dialect Society, and many anonymous reviewers who have provided commentary throughout the process.

Thanks are also due to the many people who sparked my interest in linguistics and advised me before graduate school. Thanks to Neil Nichols, who may have provided what was my first exposure to linguistics in Theory of Knowledge class at Norcross High School. In my undergraduate career, I was fortunate to have been advised by several faculty at the University of North Carolina at Chapel Hill including Elliott Moreton, Jennifer Smith, and Katya Pertsova. Elliott advised my first linguistics research project and I am very grateful for his mentorship.

There are many people outside of the linguistics community who have contributed to this process. Thanks to Gabriel Hall for your unwavering support and encouragement. Thanks also to the Western Massachusetts Sacred Harp Community, Common Ground at First Churches Northampton, and St. John's Episcopal Church for providing spiritual and musical homes during my graduate school career.

Finally, I would like to thank my family: David, Tammy, Gray, and Noelle. They have continually supported me (and put up with many bizarre conversations about language). Thanks especially to my parents, David and Tammy Hauser, for encouraging me to pursue my academic goals in the first place. I dedicate this dissertation to yall.

ABSTRACT

EFFECTS OF PHONOLOGICAL CONTRAST ON WITHIN-CATEGORY PHONETIC VARIATION

SEPTEMBER 2019

IVY HAUSER

B.A., UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Kristine Yu and Professor John Kingston

This dissertation investigates an often assumed hypothesis in phonetics and phonology: that there should be relatively less within-category phonetic variation in production in languages which have relatively more phonological contrasts (Lindblom, 1986, on vowels). Although this hypothesis is intuitive, there is little existing evidence to support the claim and it is difficult to generalize outside of vowels. In this dissertation, I argue that this hypothesis is not trivially true and needs additional specification. I propose an extension of this hypothesis, Contrast-Dependent Variation, which predicts relative differences in extent of within-category variation between languages and individual speakers. Contrast-Dependent Variation can make predictions across multiple phonetic spaces as it considers individual phonetic dimensions to be the relevant units of comparison, rather than phonological inventory subsets (stops, vowels, etc.). I therefore predict that relative differences in extent

of within-category variation can be predicted by differences in cue weight, rather than differences in number of phonemes. The dissertation tests this hypothesis by examining two between-language case studies: stops in Hindi and English and sibilants in French and Polish. I also consider a within-language case study: individual differences in extent of within-category variation in Mandarin sibilants. The results here show that differences in extent of variability between languages and speakers are systematic; they are structured according to the system of phonological contrasts.

The first between-language case study is on Hindi and English. Hindi has four stop categories at each place of articulation while English has two. Contrast-Dependent Variation predicts less variation in Hindi, but only along the particular phonetic dimensions that realize additional contrasts relative to English. This was observed in the results: Hindi speakers exhibited less variation in closure voicing both within and between speakers relative to English, but both languages exhibited similar amounts of within-category variation in voiceless lag time. I analyze multiple sources of variation in the closure voicing data in both languages. The findings support Contrast-Dependent Variation, but also have implications for feature representations in phonological theory and theories of transfer in L2 acquisition.

In the second between-language case study, I consider sibilants in French and Polish. The results demonstrate the importance of considering phonetic dimensions rather than inventory subsets in evaluating the between-language predictions of Contrast-Dependent Variation. Polish has three voiceless sibilants which contrast in place of articulation while French has two. The French sibilants are contrasted primarily in spectral center of gravity (COG). The Polish sibilants employ both COG and the second formant of the following vowel (F2) to make the 3-way place contrast. We might expect less variation in Polish along the dimensions which realize additional contrasts relative to French, in this case F2. However, F2 is used as a primary cue to vowel contrasts in French, therefore Contrast-Dependent Variation does not predict

any differences in extent of F2 variation between the two languages. In accordance with the hypothesis, no significant differences were observed in the production study.

In the within-language case study, I focus on the three-way sibilant contrast in Mandarin. There are individual differences in the phonetic implementation of the contrast, with some speakers relying more on spectral center of gravity (COG) to distinguish the sibilants while others rely on a combination of COG and the second formant of the following vowel (F2). Contrast-Dependent Variation predicts more variation in F2 in speakers that utilize the COG dimension more for contrast. This relationship is seen across speakers for the alveolar and alveopalatal sibilants.

The last section of the dissertation explores metrics for quantifying contrast in phonological inventories, considering the notions of dispersion and separability. I propose a new metric to calculate acoustic dispersion with modeled articulatory-acoustic data (Schwartz et al., 2012) as a test case. Model comparison shows that results crucially depend on metric choice.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

This dissertation examines the relationship between phonological contrast and phonetic variation. Specifically, how does the presence of phonological contrast in a given phonetic space affect within-category variation of phonetic cues in production? A major goal of this dissertation is to contribute to understanding the typology of variation: How do amounts and sources of phonetic variation vary across languages?

There is a growing body of work showing that although phonetic realization is variable, this variation is not random; it is often structured according to various non-contrastive factors.[1] The results in this dissertation show that structured non-contrastive variation does not emerge identically across all languages: phonological contrast is one mechanism which constrains the space of possible phonetic variation, resulting in predictable cross-linguistic differences in patterns of variability. In addition, despite the ubiquity of (structured and unstructured) phonetic variation, the results here also show some cases with high degrees of consistency in phonetic realization.

(Lindblom, 1986, p. 33) proposes an intuitive hypothesis about the relationship between phonological contrast and phonetic variation in vowel inventories: "the phonetic values of vowel phonemes should exhibit more variation in small than in large systems." Under this hypothesis, distributions must be tightened in a more crowded space in order to avoid overlap between categories and preserve perceptual distinc-

[1] See §1.1 for a definition of how the term *contrast* is being used in this dissertation and further discussion on terminology.

tion. A language with relatively fewer categories exploiting a single phonetic space has room for within-category variation while maintaining separation between categories. The prediction arises from the assumption that speakers aid listener perception by producing speech sounds that are sufficiently (but not maximally) perceptually distinct.

There is scant and conflicting evidence in favor of this hypothesis, which has only been explicitly tested in vowel inventories (see §1.4 for a review). However, a relationship between contrast and variation is often assumed in phonetic literature. In this dissertation, I generalize the original Lindblom (1968) hypothesis outside of vowels and examine the relationship between phonological contrast and extent of within-category variation in multiple consonant spaces. I show that the assumption of more variation in the absence of contrast, while mathematically intuitive, is not trivially true. I propose a more explicit hypothesis which I call *Contrast-Dependent Variation*: Acoustic realizations of speech sounds should exhibit less variation along a particular phonetic dimension in languages that realize phonemic contrast(s) along that dimension, relative to languages which do not realize phonemic contrasts along that dimension. This dissertation tests this hypothesis with multiple case studies using data from five languages.

In this chapter I define relevant terms (§1.1), outline the goals and hypotheses of the dissertation (§1.2), review relevant background literature (§1.3-1.4), and discuss the contributions of this dissertation (§1.5).

1.1 Definitions of terms

Throughout this dissertation, I employ the use of several terms (like "variation") which, while common, are often not clearly defined in the literature. In this section, I briefly clarify the definitions assumed in this dissertation.

1.1.1 Phonological contrast and categories

The term *contrast* is generally used to describe "a situation in which phonetic differences reflect and represent categorical differences in meaning" (Scobbie, 2006, p. 89). In this dissertation, when I use the term *contrast* I am referring specifically to *phonemic contrast* unless stated otherwise. I use the term *non-contrastive* to describe phonetic differences which do not reflect categorical differences in meaning for native speakers (i.e., they are not phonemic). Following this, a *phonological category* refers to a group of tokens (either observed or abstract) where the phonetic differences among tokens are non-contrastive. Phonemic or contrastive differences refer to phonetic differences between phonological categories.

It is not always clear whether certain phonetic differences reflect a phonemic contrast. The term "quasi-phonemic contrast" has been used to describe these cases (Harris, 1990; Hualde, 2004; Scobbie, 2006). The cases in this dissertation deal primarily with examples of phonemic contrast. The issue of quasi-phonemic contrast is further discussed in relation to category mergers in Chapters 3-4.

1.1.2 Variability, variation, variance

Throughout this dissertation, *variability* and *variation* are both used to refer to "fluctuations within a single measure, specifically within-category acoustic phonetic variability" (Vaughn et al., 2018). *Variance* is the quantitative measure of standard deviation squared, which is used to calculate variation along a particular phonetic dimension. In this dissertation, I frequently employ the coefficient of variation, a quantitative measure where variance is divided by the category mean. This allows for comparison of variance across different mean values.

There have been some attempts to draw a terminological distinction between *variation* and *variability*, where *variation* refers to fluctuation due to conditioning factors (i.e., allophonic variation) (Vaughn et al., 2018). This distinction is acknowledged

to be imperfect as not all fluctuation in phonetic realization can be easily characterized as "conditioned" or "unconditioned". For example, seemingly random variability might be due to conditioning factors which have not yet been analyzed. For this reason, I collapse this potential distinction when discussing *variability* and *variation* and will use the terms interchangeably in this dissertation.

I draw a necessary distinction between *within-speaker* and *between-speaker* variation as well as *within-category* and *between-category* variation. I also distinguish between *group-level* measures of variation and *speaker-level* measures of variation. *Within-speaker variation* refers to the variation exhibited by a single speaker where variance is calculated over a set of phonetic observations from that speaker. In this dissertation, *within-speaker variation* is always calculated over a single phonetic dimension.[2] *Group-level within-speaker variation* refers to the "extent to which members of a group are internally variable" (Vaughn et al., 2018). This can be calculated by examining within-speaker variation and then deriving an aggregate measure of those values. *(Group-level) between-speaker variation* refers to the variation exhibited among multiple speakers where variance is calculated over a set of means[3] which are calculated from the sets of phonetic observations from each individual speaker.

Similarly, *within-category variation* refers to the variation exhibited in realization of a single phonological category where variance is calculated over a set of phonetic observations along a single phonetic dimension from realizations of that particular category. These calculations may be restricted to observations from a certain speaker, certain vowel context, etc. and are then termed *within-category within-speaker variation*, *within-category within-vowel variation*, etc. *Between-category variation* refers to

[2]It would be possible to create a measure of within-speaker variation where variance is calculated over multiple phonetic dimensions to create an aggregate measure of variability for a single speaker. When using the term here, I am referring to within-category variation along a single phonetic dimension. Comparing aggregate variability across multiple dimensions is an area for future work.

[3]In these definitions, I use mean as the default measure of central tendency.

the difference in mean values across realizations of multiple phonological categories where variance is calculated over that set of mean values. These calculations may also be restricted to observations from a particular speaker or vowel context and would be termed *between-category within-speaker variation* or *between-category within-vowel variation* accordingly.

1.1.3 Phonetic dimensions, cues, and correlates

For the discussion in this dissertation, I use the term *phonetic dimension* to refer to any measure that can be extracted from the acoustic signal.[4] These can be temporal measures (e.g., voice onset time, vowel duration), spectral measures (e.g., center of gravity or formant values), or other acoustic measures. I also use the term *phonetic dimension* to refer to derived measures which have been calculated from acoustic measures such as locus equations (Sussman et al., 1991), or formant difference measurements for analyzing vowel transitions.

I use the term *correlate* to refer to any phonetic dimension which correlates with realizations of different phonological categories and the term *cue* to refer to correlates which have been shown to be relevant for perception, following Raphael (2005), among others.[5] The term *cue* can be used to refer to the relevant phonetic dimension in production or perception. As this dissertation focuses on within-category variability in production, I will typically use cue to refer to a particular phonetic dimension in production unless otherwise specified.

---

[4]This is not meant to dismiss the relevance of articulatory or other dimensions. I am using the term *phonetic dimension* as a shorthand for *acoustic phonetic dimension* because this dissertation deals with acoustic data.

[5]The relationship between acoustic correlates and perceptual cues is not necessarily one-to-one. For example, multiple correlates can integrate to create a singular perceptual effect (Repp et al., 1978; Summerfield, 1979; Kingston, 1992, among others). However, in this dissertation I focus on the behavior of cues in production. I use the term cue to refer to particular phonetic dimensions which have been shown to be relevant for perception.

Cue weight refers to a quantitative measure of relative cue strength in production or perception. In this dissertation, I will by default use the term to indicate cue weight in production unless otherwise specified. Weighting cues in production data is frequently done by applying a classification algorithm (e.g., discriminant analysis, logistic regression) where the relevant cues are predictors (Shultz et al., 2012; Garellek and White, 2015; Schertz et al., 2015; Kim and Clayards, 2019). Strength of each predictor is taken to be a metric of cue weight.[6] When discussing differences in amount of contrast or degree of contrast on a particular dimension, I am referring to relative differences in the cue weights of that dimension.

I use the term *primary cue* to refer to the cue with the highest relative weight in production and *secondary cue* to refer to cues with lower weights than the primary cue. As the studies in this dissertation are production experiments, I will use these definitions unless otherwise specified to be referring to a perceptual cue. When discussing perception, the term *primary cue* refers to the cue which has been shown in perception experiments to exert the strongest influence on perception of a given contrast.

I use the terms *separability/separable* and *dispersion/dispersed* to refer to properties of multiple phonological category distributions in phonetic space. Category *separability* refers to the degree to which tokens from two or more phonological categories overlap in the phonetic space. *Dispersion* refers to category spread in phonetic space. Categories which are dispersed are often also separable, but this is not always the case.

---

[6]Specifics of how such methods are implemented are discussed further in Chapters 4-5.

1.2 Goals and hypotheses

The main goals of the dissertation are as follows: to provide new data examining within-category variation of multiple phonetic cues in consonant spaces, to situate these findings within the current literature on phonetic variation, speech perception/production, and Dispersion Theory, and to propose testable hypotheses about the relationship between phonological contrast and the extent of phonetic variation.

The broad claim in this dissertation is that phonological contrast constrains the space of possible phonetic variation, resulting in predictable differences in patterns of variability within and between languages. The experiments that follow test two predictions made by the Contrast-Dependent Variation hypothesis. I test the between-language prediction of the hypothesis using stops in Hindi and English and sibilant fricatives in Polish and French and the within-language prediction of the hypothesis using sibilant fricatives in Mandarin.

1. Between-languages: For a given phonetic dimension X (e.g. voice onset time, spectral center of gravity), we expect less group-level within-speaker variability and less between-speaker variability in languages which employ X as a primary cue to a phonological contrast relative to languages which do not employ X as a primary cue to a phonological contrast.

2. Within-languages: Given a phonological contrast with two phonetic dimensions X and Y serving as cues and between-speaker variation in which dimension is used as the primary cue, we expect relatively more within-category within-speaker variability in X for speakers who show relatively more contrast on Y. In other words, we expect variability on X and degree of contrast on Y to be positively correlated between speakers.

The following sections of Chapter 1 present a review of the relevant background literature. In §1.3, I discuss the literature on variation in production with a focus

on sources of variation relevant to the results in this dissertation. In §1.4, I discuss the literature on Dispersion Theory with a focus on the Lindblom (1968) hypothesis about within-category variation and dispersion in consonant inventories.

1.3 Background: Phonetic variation

It is well-established that phonetic realization of phonological categories is variable both within and between speakers (Hillenbrand et al., 1995; Newman et al., 2001, among others). There is a large body of work on sources of variation in speech production, which include speaking rate, phonetic context, and sociocultural factors (to name only a few). While sources of variation are relatively well-studied, there is less work on what factors condition differences in extent of variation. The work in this dissertation builds on the current literature on variation by examining differences in extent of variation across languages and speakers.

In this section, I review the existing literature on individual differences and hyperarticulation as sources of phonetic variation in production. There are many other factors which contribute to variation in production. However, these factors are the most relevant to the results presented here. The experiments in this dissertation were laboratory studies where factors which would be sources of variation in natural speech (phonetic context, lexical frequency, etc.) were controlled. Further discussion of sources of variation in laboratory experiments is included in Chapters 2-4.

1.3.1 Individual differences

Individual speaker differences are well-documented in many phonetic spaces including differences in vowel formant frequencies among native speakers (Johnson et al., 1993; Wright, 2004; Ferguson and Kewley-Port, 2007) and L2 learners (Baker and Trofimovich, 2006), voice onset time (Allen et al., 2003; Scobbie, 2006; Theodore et al., 2009; Chodroff and Wilson, 2017), and sibilant center of gravity (Newman

et al., 2001; Tabain, 2001), among others. In some cases, the phonetic values for a particular category produced by one speaker may be almost entirely overlapping with values from the same category produced by another speaker (Newman et al., 2001 for sibilant fricatives in English; Hillenbrand et al., 1995 for vowels in English).

A growing body of work demonstrates that although phonetic values from different speakers can show a great deal of overlap, individual variation is often systematic across contrasts and phonetic dimensions. In vowels, correlations across vowel categories between talker-specific formant values have been observed (Nearey, 1989; Rose, 2010). In stops, Chodroff et al. (2015) observed correlations in mean VOT values within-speakers across different stop categories of English. Bang and Clayards (2016) builds on this, examining correlations between phonetic values of stops and fricatives. They observed correlations in VOT values among stops for individual talkers and also observed correlations between VOT and fricative duration within talkers. Clayards (2018) examined individual talker and token variation in three cues to stop voicing in English, and did not observe consistent covariation between cues. Clayards argues that this variation is structured by individual speaking styles as covaration between cues is not systematic across speakers.

This dissertation builds on this work by testing additional hypotheses about how between-speaker variation might be structured. The within-language predictions of Contrast-Dependent Variation make predictions about patterns in individual differences, proposing additional systematicity in individual differences. The previous work summarized here showed that individual differences in phonetic values are often systematic across different phonological categories. The experiment in Chapter 4 of this dissertation builds on this, showing that individual differences in extent of variation are also systematic across speakers.

1.3.2 Hyperarticulation

Speakers adopt the use of clear speech in a variety of contexts. Most important to the experiments in this dissertation is the use of clear speech in lab contexts. Differences in lab speech and spontaneous speech are well documented, including a tendency for hyperarticulation by default (Summers et al., 1988; Harnsberger et al., 2008, though see Xu, 2010).

The use of clear speech or hyperarticulation generally involves a decrease in speaking rate, an increased pitch range, and increased acoustic distance between contrasting segments (Picheny et al., 1986; Bradlow and Bent, 2002; Smiljanić and Bradlow, 2005). The increased acoustic distance between contrasting segments can affect various acoustic cues depending on the contrast. For example, in English clear speech VOT increases for voiceless stops but does not change for voiced stops (Chen, 1980; Picheny et al., 1986; Ohala, 1994; Krause and Braida, 2004). However, speakers may use different strategies to enhance contrast leading to between- and within-speaker variability even in clear speech situations (Warner and Tucker, 2011).

There has been considerably less work on extent of variation in clear speech. In this dissertation, I present results of multiple laboratory studies (where people frequently tend towards clear speech) where differences in variation are present between languages and speakers. These experiments show that it is not the case that speakers always minimize within-category variation in clear speech contexts. Further discussion of hyperarticulation effects in the particular case studies here is included in Chapters 2-4.

1.4 Background: Dispersion Theory

Dispersion Theory (c.f. Liljencrants and Lindblom, 1972; Lindblom, 1986; Schwartz et al., 1997) was originally formulated to make predictions about the relative typological frequency of vowel inventories cross-linguistically. The intuition behind the

proposal is that vowel spaces are optimized to aid in perceptual distinction. Liljencrants and Lindblom (1972) propose maximal contrast as an organizing principle in vowel inventories. They define the vowel space using two phonetic dimensions: F1 and F2′, which is a combination of F2 and F3. Their model correctly predicted the cross-linguistic frequency of /i a u/ for three-vowel inventories but had more discrepancies with predictions for larger inventories.

Several updates have been made to the original formulation of DT to address these and other discrepancies in the predictions. Lindblom (1986) added multiple revisions including the concept of sufficient instead of maximal dispersion. Dispersion from sufficient contrast predicts that languages with more phonological categories in a given space should have an overall larger phonetic space and have tighter categories within that space. For example, the [i] from a 14 vowel inventory should have lower F1 and higher F2 than the [i] from a 3 vowel inventory (on average, and scaled for speaker differences). The realizations of [i] should also show less variation in the 14 vowel inventory than in the three vowel inventory.

In this dissertation, I propose a revision of the hypothesis about within-category variation in Lindblom (1986). The original hypothesis was formulated only with respect to vowels, though it is often assumed to be true for vowels and consonants. The revision proposed here makes the hypothesis more explicit so it can be tested in phonetic spaces other than vowel inventories.

1.4.1 Previous work on the predictions of Lindblom (1986)

This section reviews the literature on the prediction about size of the phonetic space. The prediction has been examined in literature on vowels (using F1 and F2 as the relevant phonetic dimensions) and tones (F0 as relevant dimension) and the results are mixed. These studies typically have not examined the related prediction about within-category variation. The work in this dissertation aims to fill that gap

by providing comparative case studies of within-category variation across consonant inventories with different numbers of phonemic distinctions.

The prediction that larger vowel inventories should occupy larger phonetic spaces is supported by data from studies including comparisons between German (14 vowels) and Greek (5) (Jongman et al., 1989) and English (11) and Spanish (5) (Bradlow, 1995). However, other studies of vowels and inventory size have not observed the dispersion prediction. Gendrot et al. (2007) compared the vowel spaces of eight languages with inventories of different sizes and found that larger inventories did not have relatively expanded vowel spaces. Livijn (2000) compared 28 languages and found that languages with 4-8 vowels have comparably sized phonetic spaces and space only increases with 11 or more vowels.

Specific investigations of the corresponding variation prediction have been limited. Many of the studies mentioned here report variance information but do not provide a comparative analysis so it is difficult to say whether the effect was observed. Bradlow (1995) did examine variability in English and Spanish vowels and did not show differences in extent of within-category variation between the two languages.

The evidence from the literature on tone systems also provides mixed results for the predictions of DT. While earlier work (Maddieson, 1977) showed the predictions of dispersion to be observed in tone systems (larger F0 space taken up by larger tone systems), more recent work presents some contradictory evidence. Alexander (2010) compared the tone spaces of five languages with different tone inventories and the results were not in accordance with the DT predictions. She found that tone space size differed as a function of type of tone language and that level-tone and contour-tone systems may not be comparable based on number of tones. A follow-up study examined tone inventories in an additional space of onset F0 $\times$ offglide F0 and the results were again not in accordance with the DT prediction that larger inventories

should use larger phonetic spaces. To my knowledge, no studies have examined the within-category variation prediction of DT in tone inventories.

1.4.2 Dispersion Theory and consonants

DT was originally formulated to make predictions about vowel inventories (Liljencrants and Lindblom, 1972). However, the issue of whether consonant inventories are also dispersed according to similar metrics has been relatively unexplored. If speakers are aiding listener perception by constraining variation in crowded phonetic spaces, we would expect to see the prediction hold for all types of speech sounds, including consonants. This dissertation addresses the question of whether the within-category variation prediction of DT applies across different types of consonant inventories.

Most of the literature on typological frequency of consonant inventories has revolved around maximal use of available features, proposed as an organizing principle by Ohala (1979). Maximal use of available features has been formalized with Feature Economy (Clements, 2003) which describes the tendency in phonological inventories to maximize the ratio of phonological features to phonemes. The economy model does predict the ubiquity of the typologically common /bilabial-coronal-velar/ inventory in actual and randomly generated inventories (Mackie and Mielke, 2011). However, it is unclear why it should be the case that feature economy applies to consonant inventories yet phonetic dispersion applies to vowel inventories.

The principle of feature economy differs from the principle of maximizing perceptual distinction in DT. Ohala claims that maximizing perceptual distinction would result in consonant inventories like [ɗ kʼ tsʼ ɬ m r ɻ], which do not actually exist. This claim is countered by Lindblom (1986), who proposed the distinction between maximal contrast and sufficient contrast. Lindblom suggests that it need not necessarily be the case that vowel and consonant systems are organized by different principles when considering sufficient instead of maximal contrast. In a further elaboration of

the idea of sufficient contrast as a organizing principle in consonant inventories, Lindblom and Maddieson (1988) propose a relationship between consonant inventory size and complexity of consonant articulation. They divide consonants into three sets: basic articulations, elaborated articulation, and complex articulations (combinations of elaborated articulations). These sets are proposed to correlate with inventory size; smaller inventories typically only use basic articulations, larger inventories make use of the elaborated and complex articulations.

Despite the focus on alternative organizing principles, some work has shown evidence for acoustic dispersion in consonant inventoires. Boersma and Hamann (2008) propose a framework in which dispersion is emergent in sibilant systems by employing a bidirectional phonetic cue constraint model. They show that when production and perception are modeled with bidirectional phonetic cue constraints, dispersion is emergent without constraints demanding dispersion.

Although the model in Boersma and Hamann (2008) does cause emergent dispersion of sibilant inventories, it is unclear exactly what part of the model causes this (bidirectionality, phonetic constraints, the combination of the two). It is also unclear exactly how prevalent consonant dispersion patterns are typologically. Their focus is on sibilant inventories, but stop and nasal inventories are referenced as well (although no typological data is provided). They note that this type of change towards a more dispersed inventory has been observed in real diachronic change between Medieval and present Polish sibilants.

The work in this dissertation builds on this modeling work by examining dispersion in the phonetic realization of present-day Polish sibilants. The model in Boersma and Hamann (2008) simulates the emergence of dispersion between three sibilant categories across the spectral center of gravity dimension in Polish. However, the three sibilant categories are not contrasted solely along the center of gravity dimension

in the data here. Further discussion of Boersma and Hamann (2008) is included in the Polish and French case study in Chapter 3.

Other work on consonant dispersion includes Schwartz et al. (2012) who addressed the question of whether stop inventories are dispersed using data generated from a vocal tract model. They claim that the typologically common stop consonant inventory /b d g/ should be viewed as a perceptually optimal and dispersed structure just like the typologically common vowel inventory /i a u/. However, their results from analysis of 50,000 stop tokens generated by a vocal tract model show that it is not the most dispersed inventory when all places of articulation are considered.

To account for the fact that the vocal tract model results show that /b d g/ is not the most acoustically dispersed inventory, they argue that the phonetic space which is considered is not the relevant space for considering stop dispersion. The space which should be considered is modulated by articulatory considerations, namely Frame-Content Theory (MacNeilage, 1998), which is used to exclude pharyngeal and epiglottal stops from the space considered for dispersion.

The theory claims that the emergence of proto-syllables in linguistic evolution and in child linguistic development happens with articulatory exploration from mandible movements through jaw cycles with high and low points in the jaw cycle corresponding to the closed and open oral cavity. Proto-consonants emerged from the upward movements of the mandible and proto-vowels from the downward movements, thus creating syllables which show contrast between high and low points in the jaw cycle. In Frame-Content theory, all proto-consonants have upward mandible movement. Because pharyngeals and epiglottals are articulated with downward mandible movements they are excluded from the space of this articulatory exploration. The movement from a pharyngeal articulation to a vowel does not come from the upward and downward movement of the mandible as in the articulation of [ba].

In the phonetic space of articulatory exploration which excludes the pharyngeals and epiglottals, the /b d g/ configuration is the most dispersed three stop system (according to the metric used by Schwartz et al., presumably visual dispersedness). By restricting the $< F1, F2, F3 >$ space considered for dispersion to exclude pharyngeals and epiglottals, Schwartz et al. (2012) are able to revive the dispersion account as the major factor contributing to stop system organization. I return to these results in Chapter 5, where I evaluate this analysis and propose a new metric for calculating acoustic dispersion, testing it on the same data used by Schwartz et al.

1.4.3 Phonetic spaces in Dispersion Theory

Most work on DT carries implicit assumptions about the relevant space for understanding dispersion. The space for analysis of often assumed to be a subset of the entire phonemic inventory defined by a shared phonological feature. For example, work on consonant dispersion looks for dispersion within consonant inventories (rather than, for example, between consonants and vowels). The spaces in which dispersion is examined are often subsets of the consonant inventory as in Boersma and Hamann (2008) with voiceless sibilant fricatives and Schwartz et al. (2012) with voiced stops. These analyses only examine dispersion among segments in a particular subset of interest. As with the vowel inventories, these subsets are defined (either implicitly or explicitly) by phonological features to refer to particular segment classes like sibilants or stops.

The approach taken in this dissertation differs from these previous approaches as relevant spaces are defined according to phonetic dimensions, not phonological features. The predictions of Contrast-Dependent Variation crucially refer to phonetic dimensions instead a (potentially ad-hoc) subset of the phonemic inventory. I will refer to inventory subsets in discussion of the problems, hypotheses, and implications, but the hypotheses I test do not require an a priori selection of relevant phonemes.

I also do not assume that the appropriate phonetic spaces for consonants are necessarily different from those of vowels. The focus on phonetic dimensions allows for investigation of variation along any explicitly defined dimension regardless of whether that dimension is relevant for production/perception of consonants or vowels.

For example, in the case study of Hindi and English stops, defining the relevant system as the stop inventory would generally predict more variation in English relative to Hindi. As Contrast-Dependent Variation is implemented over phonetic dimensions instead of inventories, the hypothesis makes different predictions for each phonetic dimension. Understanding phonetic dimensions as the relevant spaces for evaluating Contrast-Dependent Variation makes crucially different predictions from previous analyses in DT. Further discussion of this and other examples are included for the individual case studies in each chapter of the dissertation.

1.5 Contributions of this dissertation

In summary, the previous literature on Dispersion Theory in consonants provides mixed evidence for acoustic dispersion as an organizing principle in consonant inventories. In addition, (to my knowledge) there have been no direct investigations of the DT prediction about relative differences in within-category variation in consonant spaces (Lindblom, 1986). The literature on phonetic variation in consonants documents many sources of variability, both within and between speakers. While there is a large body of work on sources and structure of phonetic variability, we have less understanding of the differences in *extent* of variability across speakers and languages.

This dissertation builds on work in DT by directly testing the hypothesis that the presence of more phonological contrasts results in less within-category variation and extending those predictions to consonant spaces. I propose a revision of the original hypothesis which can be tested across multiple phonetic spaces. This dissertation also builds on the literature on phonetic variability by examining phonological contrast

as a factor that conditions the extent of variability. The results here showcase that phonetic variation (both within and between speakers) is structured by another factor which has received relatively less attention in the literature: differences in extent of variability are also structured according to the system of phonological contrasts.

1.6 Outline of chapters

In the chapters that follow, I present three case studies testing the general hypothesis that variation emerges in the absence of contrast (Contrast-Dependent Variation; further specified in §1.2) and provide a methodological discussion on ways of quantifying dispersion and separability with application to stop inventories and Mandarin sibilants.

Chapter 2 presents the results of the experiment on stop consonant production in Hindi and English. The goal of the experiment is to examine language-specific differences in within-category variation across multiple phonetic dimensions according to the between-language predictions. Hindi has four contrasting stops at each place of articulation while English has two. Contrast-Dependent Variation predicts less variation in Hindi, but only along the particular phonetic dimensions that realize additional contrasts relative to English. This was observed in the results: Hindi speakers exhibited less variation in closure voicing both within and between speakers relative to English, but both languages exhibited similar amounts of within-category variation in voiceless lag time. I analyze multiple sources of variation in the closure voicing data in both languages. The findings support Contrast-Dependent Variation, but also have implications for feature representations in phonological theory and theories of transfer in L2 acquisition.

Chapter 3 presents the results of an experiment on sibilant fricative production in Polish and French. The results here demonstrate the importance of considering phonetic dimensions rather than inventory subsets in evaluating the between-language

predictions of Contrast-Dependent Variation. Polish has three voiceless sibilants which contrast in place of articulation while French has two. The French sibilants are contrasted primarily in spectral center of gravity (COG). The Polish sibilants employ both COG and the second formant of the following vowel (F2) to make the 3-way place contrast. We might expect less variation in Polish along the dimensions which realize additional contrasts relative to French, in this case F2. However, F2 is used as a primary cue to vowel contrasts in French, therefore Contrast-Dependent Variation does not predict any differences in extent of F2 variation between the two languages. The prediction simply refers to whether a particular phonetic dimension is employed as a primary cue in the language, and does not restrict this to a particular subset of phones. In accordance with the hypothesis, no significant differences were observed in the production study. I discuss implementation of Contrast-Dependent Variation predictions in light of these results and implications for sound change and perception.

Chapter 4 presents the results of an experiment on variation in production of sibilant fricatives in Mandarin. The goal of this experiment is to examine individual speaker differences to test the within-language predictions of Contrast-Dependent Variation. Mandarin (like Polish) has a sibilant contrast over three places of articulation and there is individual variation in how these contrasts are instantiated in the phonetic space. We expect to see relatively more variability on the F2 dimension in speakers that have more between-category variation on the COG dimension. This was observed as a general trend across speakers. I discuss these findings with reference to the other experiments and review implications for perception and sound change.

Chapter 5 considers metrics for evaluating separability and dispersion between phonetic categories. Although dispersion and separability are intuitively similar, I argue that they should be considered independent properties of phonological inventories. I discuss these differences with two case studies: stop inventories and Mandarin sibilants (from Chapter 4). The case studies provide examples of how utilizing dif-

ferent metrics changes results. For quantifying dispersion, I propose a new metric which incorporates within-category variation directly into the distance measurement. I discuss implications of the case studies and suggest methodological considerations in choosing metrics of separability and dispersion in future work.

Chapter 6 summarizes the findings and contributions of the dissertation. I discuss remaining questions and areas for future work.

CHAPTER 2

BETWEEN-LANGUAGE CASE STUDY: STOPS IN HINDI AND ENGLISH

2.1 Introduction

In order to test the revision of the Lindblom (1986) hypothesis, I compare within-category variation of stop consonants in Hindi and English using a speech production experiment. Hindi has four stop categories at each place of articulation (POA): voiced, voiced aspirated, voiceless, and voiceless aspirated, while English has two (consonant inventory charts shown in Tables 2.1 and 2.3). I compare the variation of voiceless lag time (positive voice onset time) and closure voicing in both languages.

To preview the results: Hindi and English long lag stops show comparable amounts of lag time variation within- and between-speakers, but English phonologically voiced stops show significantly more variation in closure voicing than Hindi voiced stops both within- and between-speakers. Structured non-contrastive patterns of variation emerge in the English voicing data, but not in Hindi. I use these results to argue for a revision of Lindblom's (1986) hypothesis (Hypothesis 1 of this dissertation): phonetic realization of phonemes in larger phonological systems should exhibit less within-category variation to avoid overlap, but only along the particular phonetic dimensions that instantiate additional contrasts.

There is a large body of work in phonetics uncovering structure in acoustic variability, showing that variation in speech production is not random (see Chapter 1 for a review). The results here add to this literature by showing that these structured patterns of non-contrastive variation do not emerge identically in all languages.

In this chapter, I examine differences in sources of conditioned variation and differences in extent of (what appears to be) unconditioned variation.[1] In the data here, non-contrastive patterns of prevoicing variation structured according to vowel context only emerge in the English data, and not in the Hindi data. I argue that these cross-linguistic differences can be predicted by the the structure of the phonological inventory, but how the phonological contrasts are implemented in phonetic space must also be considered. Specifically, we expect more variation in English because prevoicing serves as a primary phonetic correlate of a phonemic contrast in Hindi, but a secondary correlate of a contrast in English.

2.1.1 Hindi background

Hindi is one of several Indo-Aryan languages which exhibit a four-way stop contrast.[2] The four-way contrast occurs at four places of articulation: bilabial, dental, retroflex, and velar. This system is often understood as a laryngeal contrast over the two dimensions of voicing and aspiration (Dutta, 2007). Assuming binary features, fully crossing the values of voicing and aspiration results in four distinct phonological categories.

Voice onset time (VOT) has frequently been analyzed as a phonetic correlate to these stop contrasts (Lisker and Abramson, 1964; Abramson and Lisker, 1967; Poon and Mateer, 1985). VOT is a duration measure of the onset of voicing relative to the release of the stop occlusion. Historically, VOT has been implemented as a continuum of negative and positive values. Voicing before the stop closure is analyzed as negative VOT and voicing which begins after the stop closure is analyzed as positive VOT.

---

[1] We cannot say with certainty that any variation is random or unconditioned because it could always be the case that we have not analyzed the appropriate factors which condition the variation.

[2] Dutta (2007) cites UPSID (Maddieson, 1981) which contains ten languages from six families with the four-way contrast.

Table 2.1. Consonant inventory of Hindi (Ohala, 1983)

	Labial	Dental/Alveolar	Retroflex	Palatal	Velar	Glottal
Stop	p b	t d	ʈ ɖ		k g	
Aspirated stop	p^h b^h	t^h d^h	$ʈ^h$ $ɖ^h$		k^h g^h	
Affricate				tʃ dʒ		
Fricative	f v	s z		ʃ		h
Nasal	m	n		ɲ	ŋ	
Approximant		l		j		

Using only VOT to characterize stop contrasts in phonetic space has been recognized as inadequate for languages like Hindi which have stops that are produced with lead voicing and aspiration (Lisker and Abramson, 1964; Schiefer, 1986; Dixit, 1989). VOT separates only stops that differ in laryngeal timing and does not distinguish stops that incorporate another feature difference. In the case of Hindi voiced aspirated stops, width of glottal opening (Benguerel and Bhatia, 1980) has been argued to be a second necessary feature distinguishing the breathy/aspirated stops from the other stops.

Various alternatives for analyzing the voiced aspirates have been proposed and can be categorized into roughly two hypotheses: voiced aspiration is a result of two independent gestures of voicing and aspiration, and, voiced aspiration is an independent mode of phonation (Dutta, 2007). Exactly what features/phonetic dimensions distinguish the voiced and voiced aspirated stops in Hindi is not directly relevant to this study because this experiment aims to compare variation in Hindi and English. The voiced aspirates are not a focus as they do not have a correspondent in the English phonological system.

VOT has been standardly defined with negative values for lead voicing (Lisker and Abramson, 1964; Cho and Ladefoged, 1999), but the use of a single measure for lead and lag voicing has been challenged. Mikuteit and Reetz (2007) use data from East Bengali (another language with a four-way stop contrast) to argue that lead voicing and lag voicing should not be considered as part of a single continuum. They instead

Table 2.2. Feature specifications for stops in Hindi

	[-spread glottis]	[+spread glottis]
[-voice]	/t/	/t^h/
[+voice]	/d/	/d^h/

propose separate duration measures of after closure time (duration from release to onset of voicing; lag time), onset voicing (start of glottal pulsing to release in initial stops), and connection voicing (closure duration in medial stops).

Following this analysis, I consider lag time (traditionally known as positive VOT) to be a separate phonetic dimension from lead time (traditionally known as negative VOT). In all discussion that follows, I use *lag time* to indicate the duration between the stop burst and onset of voicing, CD to indicate closure duration, and CV to indicate closure voicing (see §2.5.1 for further discussion of how voicing was operationalized in the experiment).

In terms of distinctive features, Hindi is typically described as fully crossing all values of two features [± voice] and [± spread glottis] (Dutta, 2007), shown in Table 2.2. In discussion of data in the present study, I refer to instances of [+voice] as *phonologically voiced stops* and instances of [+sg] as *phonologically aspirated stops*. There is some debate in the literature about the feature specification of the voiced aspirates (Benguerel and Bhatia, 1980; Dixit, 1989; Dutta, 2007). There is also debate about whether these features should be binary or privative (Honeybone, 2005; Schwarz et al., 2019), which is not specific to Hindi. For the laryngeal relativists, the question of feature binarity is independent from the question of what the features are. The questions examined here do not hinge on any particular feature representations and I discuss implications for feature specification in §2.5.5.

2.1.2 English background

English has two contrasting stop consonants at three places of articulation: bilabial, alveolar, and velar. These can be seen in the English consonant inventory given

Table 2.3. Consonant inventory of English (Quirk et al., 1972)

	Labial	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Stop	p b		t d			k g	
Affricate				tʃ dʒ			
Fricative	f v	θ ð	s z	ʃ ʒ			h
Nasal	m		n			ŋ	
Approximant			l ɹ		j	w	

Table 2.4. Representations of English stops

Phoneme	Distinctive feature (realist)	Distinctive feature (relativist)	Common phonetic realizations in word initial position
/b/	–	[+voice]	[b p]
/p/	[sp. gl.]	[-voice]	[p^h]

in Table 2.3. In American English, lag time is the primary cue to the stop contrast and other phonetic cues such as F0 frequently co-vary with lag time (Lisker, 1986; Lisker and Abramson, 1967; Keating, 1984, among others). Because lag time is the primary cue, English is often considered an "aspirating" language instead of a "true voicing" language. Despite this, the English stops are typically represented using the IPA symbols for voiceless and voiced stops /t d/.

There is some debate in the literature about which phonological features should be used to distinguish English stops. The laryngeal realist view takes the position that the phonological features should reflect the phonetic realization (Honeybone, 2005; Jessen and Ringen, 2002; Beckman et al., 2013, among others). Under this view, the feature distinguishing the two English stops is [spread glottis] (features are also typically privative in this view). The laryngeal relativist view takes a more abstract approach focusing on cross-linguistic similarities. In this view, two-way stop contrasts are represented with [(±)voice] and phonetic implementation can differ across languages (Keating, 1984; Kingston and Diehl, 1994; Lombardi, 1994; Cyran et al., 2011, among others). The table in 2.4 shows these potential representations of the English stop phones.

In this chapter, I assume the [±voice] analysis and I revisit the question of feature representation in §2.5.5. In all discussion that follows, I refer to the English short lag stops /b d g/ as *phonologically voiced* and the English long lag stops /p t k/ as *phonologically voiceless*.

Despite the classification of English as an aspirating language, several studies have reported prevoicing on English phonologically voiced stops, which is assumed to be the primary phonetic correlate of voicing in "true voicing" languages. In work on British English, Docherty (1992) reported prevoicing (as percentage of voicing during the stop closure) from five adult male speakers of Southern British English. The mean percentages of voicing during the closure was 51% of CD for [b], 58% of CD for [d], and 66% of CD for [g]. Deterding and Nolan (2007) found similar results with seven British English speakers. Both studies elicited the stops in a word-initial utterance-medial post-vocalic context (the same context used in the present study).

In work on prevoicing in American English, Lisker and Abramson (1964) report data on utterance initial stops (n=4) and note that some English speakers produced prevoiced stops in this context. However, one speaker produced 95% of all prevoiced stops. Davidson (2016) also found prevoicing variation in connected read speech to be influenced by linguistic factors such as adjacent sounds and lexical stress. There is also a body of work documenting prevoicing in Southern American English varieties (in the utterance-initial and medial contexts), sometimes with higher incidence among male and African-American speakers (Jacewicz et al., 2009; Elston et al., 2016; Herd et al., 2016; Hunnicutt and Morris, 2016). These findings suggest that the frequency of prevoiced phonologically voiced stops may be determined by the regional dialect of English speakers.

Table 2.5. Hypotheses about within-category variation summarized

Hypothesis	**Summary**	**Domain**	**Phonetic space**
Lindblom (1986)	"The phonetic values of vowel phonemes should exhibit less variation in small systems than in large systems."	vowels	F1/F2 assumed
Contrast-dependent variation (proposed here)	For a given phonetic dimension X, we expect less group-level within-speaker variability and less between-speaker variability in languages which employ X as a primary cue to a phonological contrast relative to languages which do not employ X as a primary cue to a phonological contrast.	general	any

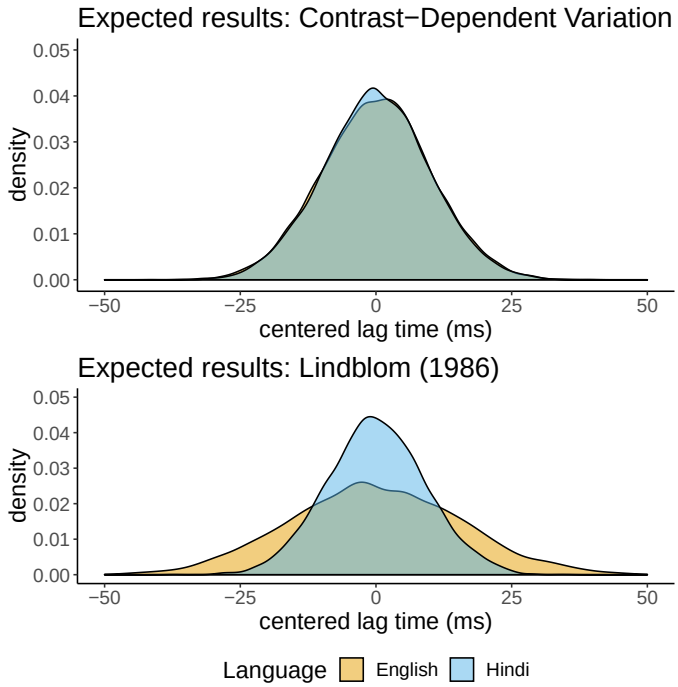
2.2 Predictions

Lindblom's (1986) hypothesis is that vowels in systems with more phonological contrasts should show less within-category variation than vowels in systems with fewer contrasts (see Chapter 1 for an extended review). In extending this prediction outside of the F1/F2 space assumed for vowels, I have operationalized a revised hypothesis, which is given in Table 2.5.

In the Dispersion Theory analyses, the relevant space is assumed to be the vowel space. While there are no explicit criteria for determining relevant phonetic dimensions, the first two formants are used in analyses. In trying to implement Lindblom's hypothesis as directly as possible, we might consider the stop inventory to be the relevant "system" as Lindblom considered the vowel inventory to be the relevant "system".

Under this assumption, Hindi stops are hypothesized to vary less relative to English stops because there are more stop phonemes in Hindi. Lindblom's prediction does not distinguish between different phonetic dimensions outside of vowels and does not address how within-category variation should be quantified. Therefore, under this

Figure 2.1. Predicted voiceless lag time distributions for two hypotheses in Hindi and English. Predictions of Contrast-Dependent Variation in top panel. Predictions of Lindblom (1986) in bottom panel.



hypothesis, one potential prediction could be that we expect voiceless aspirated stops in Hindi to vary less in lag time relative to English. Expected results under this prediction are shown in Fig. 2.1.

If we define the relevant "system" according to single phonetic dimensions instead of overall number of phonemes, we expect no difference in lag time variation (a schematic of predicted results is given in Figure 2.1). This is because because both languages use the dimension of lag time to distinguish one contrast between short lag and long lag stops (see Table 2.6). While there are four phonological contrasts in Hindi at each place of articulation, the lag time contrast is not a four-way contrast. When lead time and lag time are not considered as part of the same continuum, Hindi has a single lag time contrast between the voiceless unaspirated and voiceless aspirated stops, similar to the single contrast in English.

Table 2.6. Phonetic dimensions in Hindi and English stops

↑	← voiceless aspiration →	
duration of	/t/	/t^h/
closure voicing	/d/	/d^h/
↓	← voiced aspiration →	

← voiceless aspiration →	
/d/	/t/

Because there is a single contrast that primarily exploits the lag time dimension in both languages, we do not expect any differences in extent of lag time variation under the revised hypothesis. However, we do expect more variation in English along the voicing dimension relative to Hindi. English does not contrast any additional stops along the voicing dimension, but Hindi does. Therefore, we predict more variability in stop closure voicing in English relative to stops in Hindi.

2.3 Experimental design

2.3.1 Participants

All speakers were between the ages of 18-30 and recruited at The University of Massachusetts Amherst. Most of the English speakers were undergraduates enrolled in introductory linguistics courses and most of the Hindi speakers were master's degree students in varying fields at the university. In the first round of data collection, nine speakers of each language were recorded.

Exclusion was determined based on two factors: an expression of difficulty or discomfort with the task and a numerical cutoff of speaking rate, as measured by pauses between the carrier phrase and the stimulus. The task was a production task which involved reading phrases off a computer screen. Therefore, native speakers with poor reading skills spoke unnaturally during the task and produced many speech errors. Any participants who expressed difficulty with the task and/or paused before

the stimulus leaving silence for more than 1.5 seconds on at least 75% of the phrases were removed from the analysis.

Five Hindi speakers and one English speaker were excluded according to these criteria. Two Hindi speakers were additionally removed from the analysis because they were L2 speakers of Hindi. This was determined by their answers to a demographic questionnaire about language background. Two English speakers were additionally removed because they did not complete the task. After exclusions, data from two Hindi speakers from the first round of data collection were retained.

To replace the Hindi speakers which were excluded in the first round, we ran a second round of data collection with a few adjustments. The call for participants was circulated in Hindi orthography to ensure the participants were comfortable with reading in addition to speaking. The experimenter was a native speaker of Hindi who spoke Hindi to the participants throughout the experiment.[3] This helped in resolving confusion among the participants about L1/L2 status of Hindi before they participated. After this second round of data collection, recordings from six speakers of each language were available for analysis.

2.3.2 Stimuli

The goal was for stimuli to be as similar as possible between languages. The stimuli were C_1VC_2 words and non-words where C_1 is a stop and V is one of [i a u]. The coda consonant of the stimulus (C_2) was in most cases a stop. If there were no stops available that could make a phonotactically natural word or non-word, then a fricative was used. If there were no fricatives available, then a sonorant was used. Eliciting only monosyllabic words avoided potential problems with placement of

[3]This was the only difference in the procedure of the experiment between the first round and the second round of data collection. The speakers whose data were retained from the first round of collection did not systematically differ in extent of lag time variance relative to those in the second round of collection.

Table 2.7. Example stimuli: Type indicates word/non-word status and additional word frequency levels in English

Language	C_1	vowel	stimulus (IPA)	type
Hindi	b	i	bit	word
Hindi	k^h	i	k^hil	word
Hindi	b^h	u	b^hut	word
Hindi	d	a	dag	word
English	p	i	pis	word-hi
English	t	a	tak	non-word
English	t	u	tub	word-hi
English	b	a	bag	word-low

stress. All stimuli were recorded in a uniform carrier phrase: "Say X again" in English and "Dobara X doharao" (repeat X again) in Hindi. The carrier phrases placed the target words in focused environments in both languages. The stimuli were all developed in consultation with native speakers to assure phonotactic wellformedness. Example stimuli are given in Table 2.7.[4]

The stimuli were grouped according to different factors depending on the language. Real words and non-words were used in both the Hindi and English stimuli. The English stimuli were grouped further according to lexical statistics to allow for analysis of lexical effects (lexical statistics obtained from the English Lexicon Project; Balota et al. (2007)). The lexical statistics are not as readily available for Hindi so the English data were checked for lexical effects under the assumption that if there were relevant differences according to lexical statistics, the Hindi productions would likely differ in the same way. Hindi stimuli were crossed according to the following factors: consonant (16 levels) × vowel context (3 levels) × word status (2 levels: word/non-word) for a total of 96 distinct stimuli. English stimuli were crossed according to: consonant (6 levels) × vowel context (3 levels) × word status (4 levels: high frequency/low frequency/non-word/has C_1 minimal pair) for a total of 72 distinct stimuli.

[4]For a full list of stimuli and other experimental materials see the public archive for this dissertation at https://osf.io/2famr/.

2.3.3 Recording

The participants were all recorded in a sound-attenuated booth using Audacity software (Audacity Team, 1999-2014). The recordings were done using an M-Audio Fast Track Pro Mobile Audio Interface and a Shure SM10A head-worn microphone. The recordings were sampled at a rate of 44.1 kHz with a bit depth of 16. The participants were presented with stimuli in the relevant orthography on a laptop computer inside the booth. They were asked to produce the phrases as naturally as possible. The research assistants were trained to give feedback which encouraged natural production.[5] The stimuli were recorded in four separate blocks, each with a different random order, totaling four repetitions of each stimulus for analysis.

The recordings from each speaker were first scanned by the author and/or a native speaker research assistant for speech errors. After these exclusions, there were a total of 3663 tokens analyzed. The recordings were forced aligned using the Montreal Forced Aligner (McAuliffe et al., 2017). The forced aligner creates Praat (Boersma et al., 2001) textgrids marking boundaries at the word and segment level. A new forced aligner model was trained on the Hindi acoustic data, which can be used in future work. Additional information about forced alignment and the Hindi model is given in Appendix A.2.

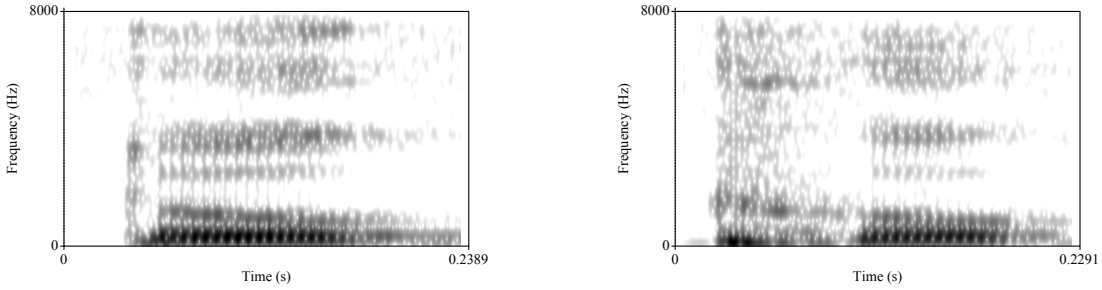
2.4 Lag time

2.4.1 Analysis: Lag time

In this section, I detail how lag time was analyzed for the phonologically voiceless stops and summarize the results. In accordance with the revised prediction, there was

[5]This included things like suggesting the participant speak as if they were talking to a friend and not giving a presentation, suggesting they say the phrase "in one breath" to discourage pausing before the stimulus, etc.

Figure 2.2. Voiceless short and long lag stops in Hindi. Left: CV sequence from token of /tup/). Right: CV sequence from token of /t^hup/).



no significant difference in group-level within-speaker lag time variability between the two languages.

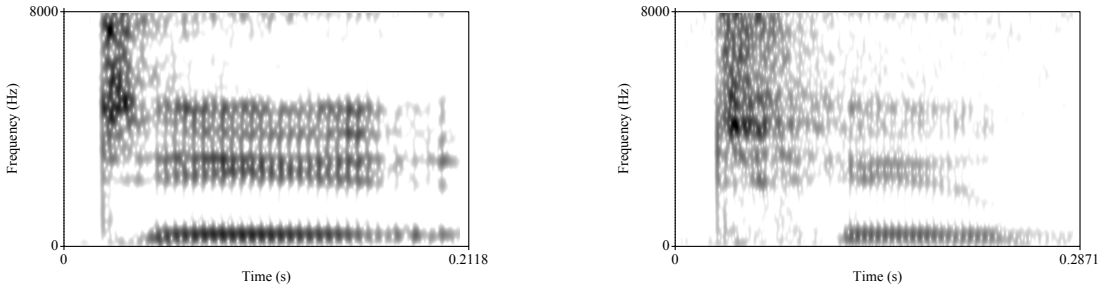
Many dialects of Hindi are currently undergoing a merger between the voiceless aspirated labial stop /p^h/ and the voiceless labiodental fricative /f/, where both are produced as /f/ (Dutta, 2007). All of the speakers in this study consistently produced the fricative, so the labial stops are not analyzed here. The coronal and velar stops in both languages are compared. The coronal category included the dental and retroflex stops in Hindi compared with the alveolar stops in English.

The forced aligned boundaries were used as input to AutoVOT (Keshet et al., 2014) which allowed for automatic measurement of lag time intervals.[6] AutoVOT was used to measure lag time for the voiceless short and long lag stops in both languages. Lag time was measured from the start of the burst to the onset of voicing. The intervals created by AutoVOT were all hand-checked and hand-corrected by the author. A random sample was additionally spot-checked by a research assistant.

Example tokens are shown in Figures 2.2-2.3. In both figures, the short lag tokens are shown on the left and the long lag tokens on the right. A difference in the duration of aspiration between the short and long lag tokens can be seen in both languages. Lag times for the long lag stops are analyzed in this section.

[6]Many thanks to Eleanor Chodroff for making her AutoVOT tutorial Praat scripts publicly available (https://www.eleanorchodroff.com/tutorial/autovot/autovot-intro.html).

Figure 2.3. Short lag and long lag stops in English. Left: CV sequence from token of /dit/. Right: CV sequence from token of /tip/.



To abstract over differences in mean values between speakers and vowel contexts, lag time values were centered around within-speaker within-category within-vowel means. A standard outlier rejection method was applied before analysis, excluding tokens with a z-score greater than the absolute value of 3 (Well and Myers, 2003). This removed 30 of 3663 tokens.

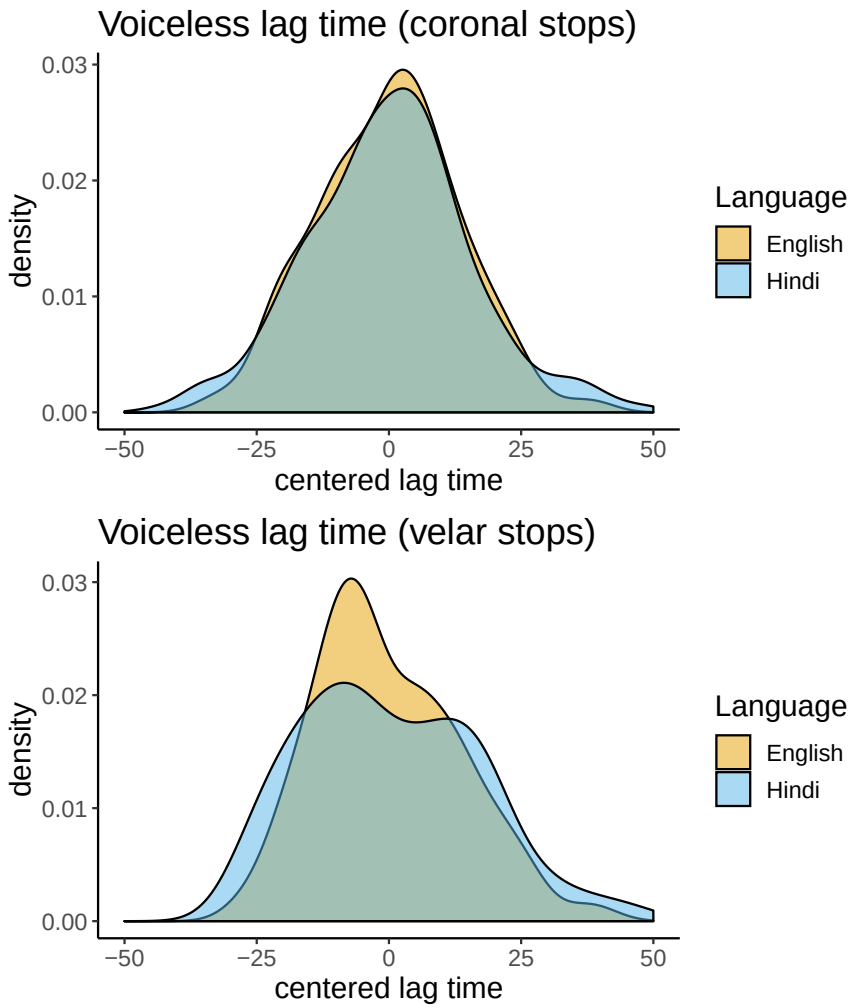
2.4.2 Results: Lag time

In Figure 2.4, I show results for long lag stops in both languages at coronal and velar places of articulation. These plots show the distributions of the centered lag time values, collapsed over speakers. Lindblom's hypothesis predicts less within-category variation in Hindi (Fig. 2.1). If this were the case, the English distributions would be wider than the Hindi distributions in the results. However, in Figure 2.4 the English distributions do not appear to be wider than the Hindi distributions for either place of articulation. In fact, it appears that Hindi speakers might actually produce more variation than the English speakers.

To quantify the differences between the languages, I use a mixed effects linear regression where within-speaker within-category lag time variance is the dependent variable.[7] Language, place of articulation, vowel (V), and their interactions were

[7]Using within-category variance as a dependent variable was also done in Vaughn et al. (2018) to test for differences in group-level within-speaker variability.

Figure 2.4. Experimental results for coronal and velar long lag stops, lag time in ms



included as predictors with random intercepts for speaker. No random intercepts for speaker were included as this additional model structure was not justified by the research question. We are mostly interested in the main effect of language, and speaker is fully nested within language. R (R Core Team, 2013) was used for all statistical analyses. The lmer function in the lme4 package (Bates et al., 2007) was used for the regression model, using LmerTest to obtain p values (Kuznetsova et al., 2017). Place was coded as a categorical variable with two levels: coronal and velar. Default dummy coding contrast structure was used with English coronal /a/ as the reference level.

Under Lindblom's hypothesis, we expect less group-level within-speaker variation in Hindi relative to English, therefore we would expect a significant effect of Language in the model. Under Contrast-Dependent Variation, the revised hypothesis I propose here, we do not expect a difference in group-level within-speaker variation, therefore we expect no significant effect of Language in the model.

Although Language is the main effect of interest, other factors were included to ensure that a significant effect of language would not be due to covariation with other factors. It is possible that stop place of articulation and vowel quality may independently influence extent of variation and therefore are included as additional factors. Random intercepts for speaker were included as the question of interest here is about differences in group-level within-category variation across the two languages. The default lmer contrast structure was used.

The model output is given in Table 2.8. In the model results, we see no significant effect of Language, in accordance with Contrast-Dependent Variation. There was also no effect of place, but there was a significant effect of the /u/ context. However, the non-significant language × /u/ interaction indicates that the vowel effect is consistent across the two languages. There are no significant differences between the

Table 2.8. Fixed effects table for linear mixed effects regression. Dependent variable: within-category within-speaker lag time variation (quantified by coefficient of variation). Predictors: language, place of articulation, V (vowel context), language × place, language × V, random intercepts for speaker. Model intercept is English coronal /a/ context. Standard error values are similar between effects when there is similar n in each level of the factor.

Fixed effects	Estimate (se)	t	p
(Intercept)	16.37(2.51)	6.53	$< 0.001^{***}$
language-Hindi	2.18(3.38)	0.64	0.53
place-velar	−0.92(1.64)	-0.56	0.580
V-/i/	−3.46(2.01)	-1.72	0.090
V-/u/	−4.44(2.01)	-2.21	0.031^{*}
language-Hindi × place-velar	−1.57(2.17)	-0.72	0.472
language-Hindi × V-/i/	1.51(2.59)	0.58	0.563
language-Hindi × V-/u/	1.87(2.59)	0.72	0.472

two languages in within-category lag time variation for either place or in either vowel context.

2.4.3 Interim discussion: Lag time

Despite the difference in number of stop contrasts in the two languages, the amount of group-level within-speaker lag time variation was similar in both languages. This is not expected under the most direct implementation of Lindblom (1986) which predicts less variation in languages with more phonological contrasts. Phonological contrasts are implemented in a multidimensional phonetic space and this prediction must be re-formulated in terms of phonetic dimensions instead of phoneme inventories.

Under Contrast-Dependent Variation (Table 2.5), similar amounts of lag time variation in Hindi and English are expected. Hindi and English both distinguish one contrast primarily using lag time when voiceless lag time is considered to be a separate dimension from prevoicing. The results here can be interpreted as providing additional evidence for the division of lag time and lead time into separate dimensions (Mikuteit and Reetz, 2007). Prevoicing and lag time pattern differently in the data here–lag time variation is similar in both languages, but closure voicing variation

differs between languages. An analysis of prevoicing and lag time which considers them as separate phonetic dimensions allows us to capture the differences observed here.

2.5 Voicing

In this section, I detail how voicing was analyzed for the phonologically voiced stops and results. The summary of results includes analysis of individual differences and vowel context effects. I also model the results using regression and compare extent and sources of variation between the two languages using model comparison. In accordance with the revised prediction, there is more variability in closure voicing in English relative to Hindi both within and between speakers.

2.5.1 Analysis: Voicing

Closure duration (CD) and closure voicing (CV) were hand measured for all stops. CD was measured from the offset of the preceding vowel until the stop burst. Tokens with stop closures longer than 300ms were excluded. CV was measured as the portion of the stop closure which contained periodicity in the waveform which indicates voicing. Tokens where periodic voicing during the closure stopped and started again were excluded (30 total tokens).

The percentage of the closure containing voicing was calculated from the measurements of CD and CV. The percentage data was also classified according to three categorical bins: no prevoicing (voicing through 0-25% of the stop closure), partial prevoicing (25-90%), and full prevoicing (90-100%). The classification of full prevoicing as voicing through 90% or more of CD follows the categorization in Beckman et al. (2013).[8]

[8]The motivation for these classifications in Beckman et al. (2013) is to only classify tokens as fully prevoiced if they are produced with active voicing instead of passive voicing. It is argued to be unlikely that passive voicing would carry on through 90% of the CD.

Figure 2.5. Voiced unaspirated and aspirated stops in Hindi. Left: CV sequence from token of /dut/. Right: CV sequence from token of /dhup/.

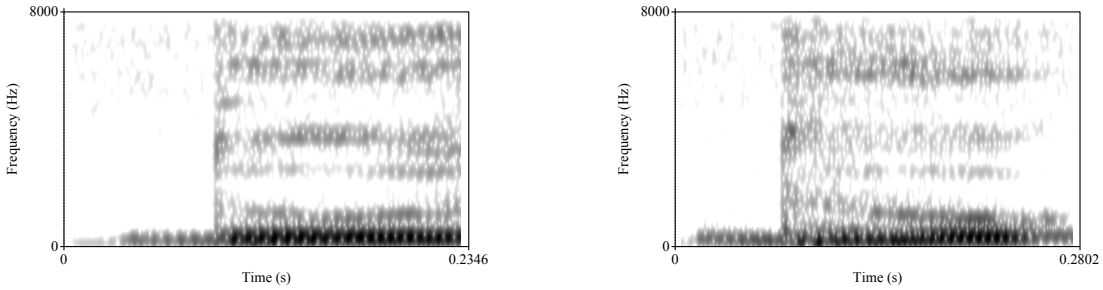
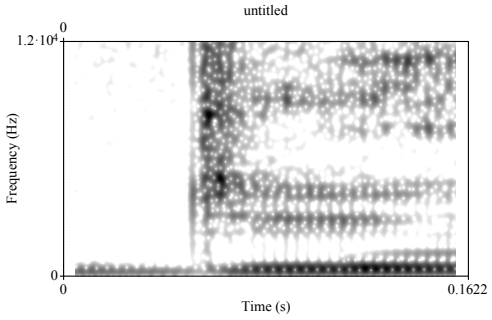


Figure 2.6. Short lag stop with closure voicing in English



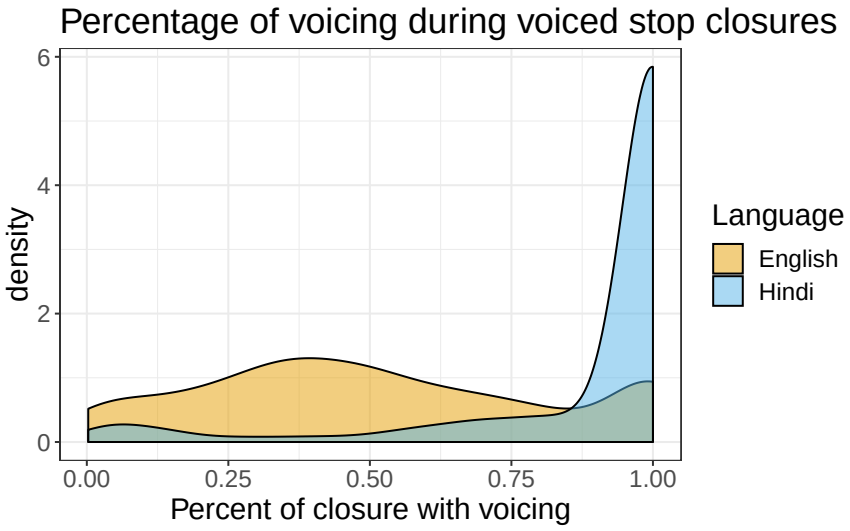
Example Hindi tokens are shown in Figures 2.5-2.6. The Hindi tokens both show voicing before the stop closure which continues through the burst into the vowel. The English token differs from the other phonologically voiced English token shown in Figure 2.3. In the token shown here in Figure 2.6, voicing starts before the stop burst. Phonetically voiced and voiceless realizations of the English [+voice] stops were observed in the data here.

2.5.2 Results: Voicing

In this section, I concentrate on within-category variation in the phonologically voiced stops in both languages.[9] Fig. 2.7 provides a density plot of the closure voicing percentages in both languages. In Hindi, the distribution of proportion voiced is

[9]Because the analysis is comparative and there are no voiced aspirate stops in English, the voiced aspirated stops have excluded from the analyses here. The results do not change (there is still more variation in English relative to Hindi) if the voiced aspirated stops in Hindi are included.

Figure 2.7. Density plot of percentage of voicing during stop closures

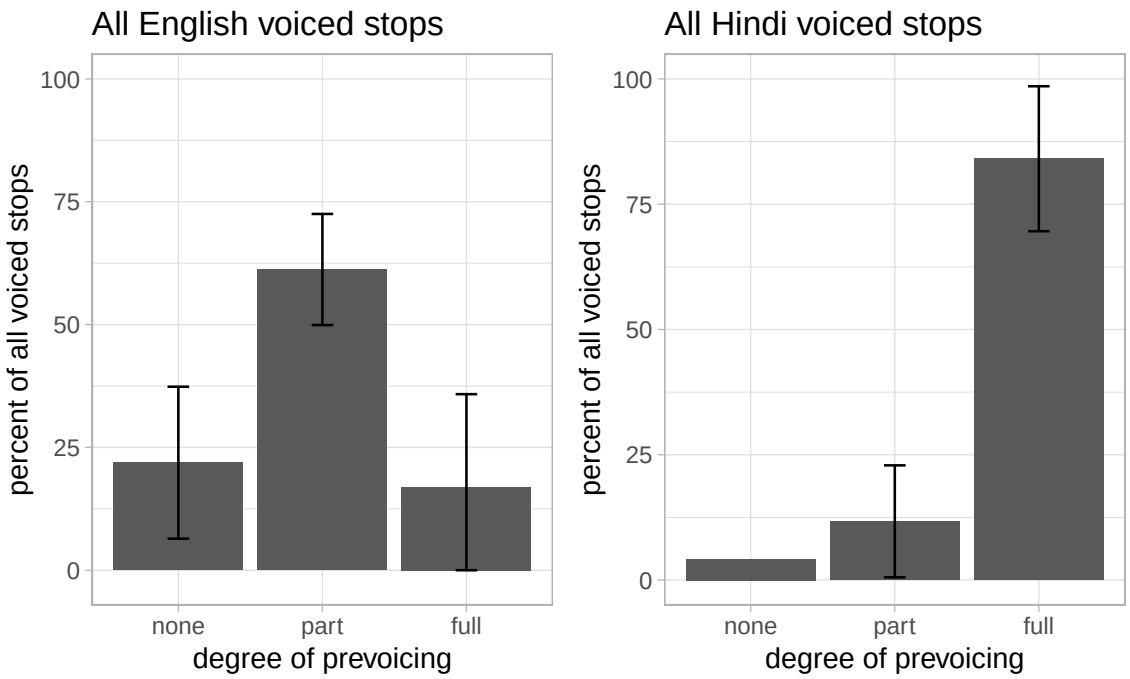


skewed as almost all stops are produced with voicing during 100% of the closure. In English, the distribution of voicing is more variable.

In Figure 2.8, I show the categorical voicing bins in both languages (no prevoicing, partial prevoicing, full prevoicing), error bars show standard deviation between speakers. In Hindi, almost all voiced stops are produced with full prevoicing (voicing through at least 90% of CD). In English, there is more overall variation in degree of prevoicing. Most of the phonologically voiced stops produced in English are partially prevoiced but this varies between speakers.

As in the lag time analysis, I use mixed effects linear regression where within-speaker within-category variation is the dependent variable. For this analysis, the dependent variable was calculated by determining the variance in closure voicing (percent of closure which has voicing) within category, speaker, and vowel context, which was then used to calculate the coefficient of variation. Language, place of articulation, vowel, and their interactions were included as predictors with random intercepts for speaker. Default dummy coding contrast structure was used with English coronal /a/ context as the reference level.

Figure 2.8. Voicing during stop closure in phonologically voiced stops (categorical bins); Error bars show standard deviation between speakers.



The model output is given in Table 2.9. Under Contrast-Dependent Variation, we expect less group-level within-speaker variation in Hindi relative to English. We therefore expect a significant effect of Language in the model, which was observed, along with several other significant effects.

The significant effect of the velar place indicates more voicing variance for velar stops. However, the negative estimate on the significant interaction between language and velar place indicates less variation for Hindi velars relative to the English coronal intercept. There were also significant main effects of the vowels /i u/ indicating less within-category variation before these vowels relative to /a/. In this case as well, an interaction term shows that this may not be the case in Hindi. The Hindi × /i/ interaction is significant with a positive intercept indicating more variation relative to the English /a/ reference level.

Table 2.9. Fixed effects table for linear mixed effects regression. Dependent variable: within-category within-speaker voicing variation (quantified by coefficient of variation). Predictors: language, place of articulation, V (vowel context), language × place, language × V, random intercepts for speaker. Model intercept is English coronal /a/ context.

Fixed effects	Estimate (se)	t	p
(Intercept)	54.19(8.32)	6.52	$< 0.001^{***}$
language-Hindi	−40.76(11.65)	-3.652	0.002^{**}
place-labial	3.82(5.89)	0.65	0.518
place-velar	16.19(5.89)	2.75	0.007^{**}
V-/i/	−16.35(5.89)	-2.77	0.006^{**}
V-/u/	−14.58(5.89)	-2.48	0.015^{*}
language-Hindi × place-labial	−1.67(7.79)	-0.21	0.831
language-Hindi × place-velar	−18.43(7.79)	-2.37	0.028^{*}
language-Hindi × V-/i/	17.31(7.79)	2.22	0.028^{*}
language-Hindi × V-/u/	10.75(7.79)	1.38	0.171

Overall, these results show that there is less group-level within-category within-speaker voicing variation in Hindi relative to English. There is less variation before /u/ in both languages relative to /a/. In English, there is also more voicing variation for velar stops and less variation before both high vowels relative to /a/.

In the next section, I examine the voicing variation further, turning to between-language differences in how the variation is structured.

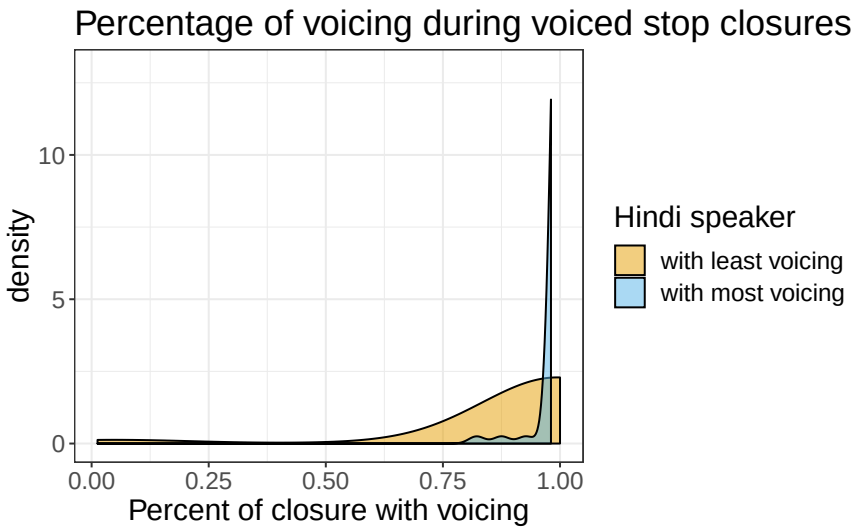
2.5.3 Results: Structure in voicing variation

2.5.3.1 Between-speaker variation

The Hindi pattern of voicing is consistent across all speakers. Fig. 2.9 shows distributions for the two Hindi speakers with the most between-speaker difference in amount of voicing. I also provide the data in terms of discrete voicing categories in Fig. 2.10. Both graphs show similar patterns between the Hindi speakers.

While all Hindi speakers consistently fully voice phonologically voiced stops, English speakers show individual preferences for degree of closure voicing. Some English speakers have voicing through 100% of the closure on almost all phonologically voiced

Figure 2.9. Hindi speakers with greatest difference in voicing (continuous)



stops while others have little closure voicing. Figure 2.11 shows the English speakers that can be characterized as having the least and most voicing during phonologically voiced stop closures. This density plot shows two distinct distributions with different means and shapes. Figure 2.12 provides the same data binned into voicing categories.

2.5.3.2 Variation across vowel contexts

Smith and Westbury (1975) reported more prevoicing in English stops before high vowels relative to low vowels. I observe a similar pattern in the English data, but not in Hindi. Just as the pattern of voicing in Hindi is consistent across speakers, the pattern of voicing is also consistent across vowel contexts. The data for both languages are shown in Figures 2.13-2.14.

2.5.3.3 Modeling sources of variance

In this section, I compare the effects of different factors in accounting for overall voicing variance in both languages. Due to the dependent variable (percentage of closure with voicing) being continuous proportion data, I use Beta Regression (Ferrari and Cribari-Neto, 2004), which is intended for proportion data bounded between (0,1). Unlike a standard linear regression which assumes the data follow Gaussian

Figure 2.10. Hindi speakers with greatest difference in voicing (categorical)

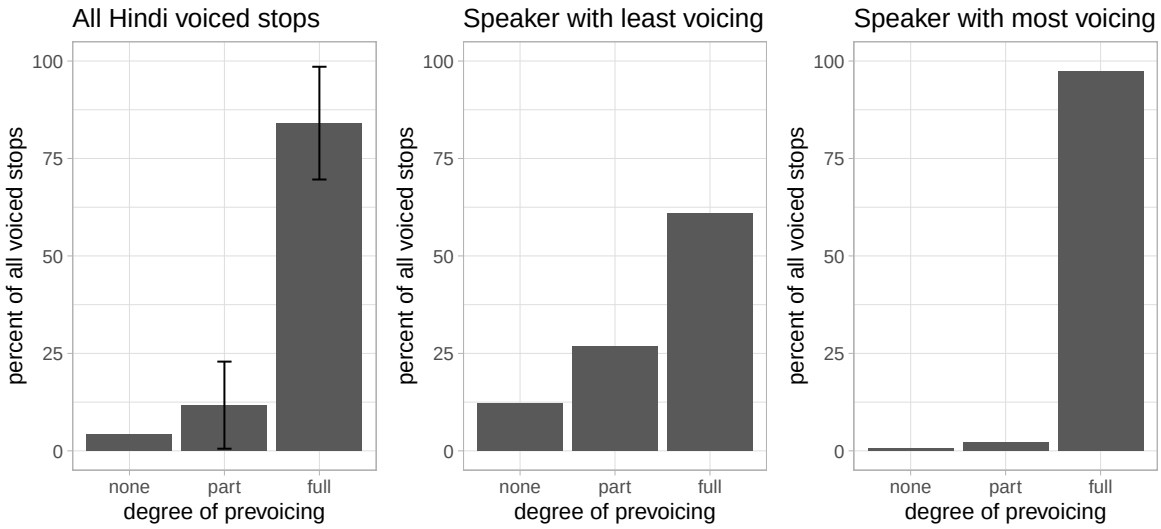


Figure 2.11. English speakers with greatest difference in voicing (continuous)

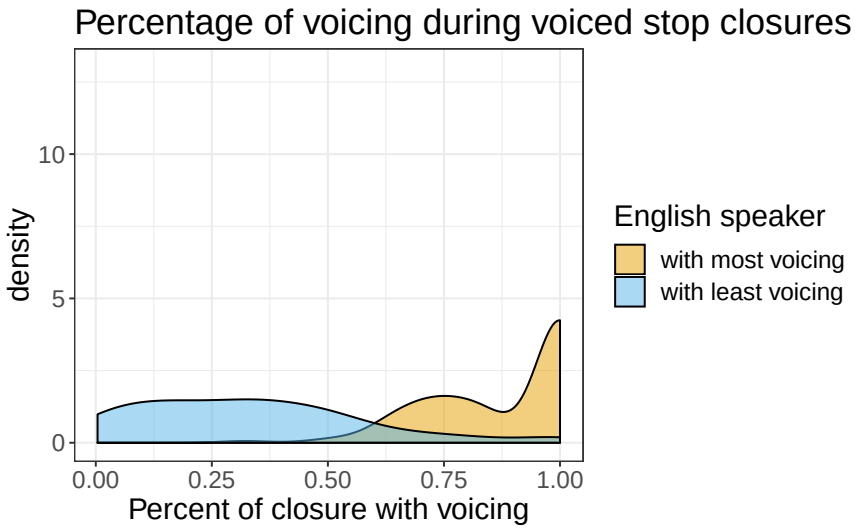


Figure 2.12. English speakers with greatest difference in voicing (categorical)

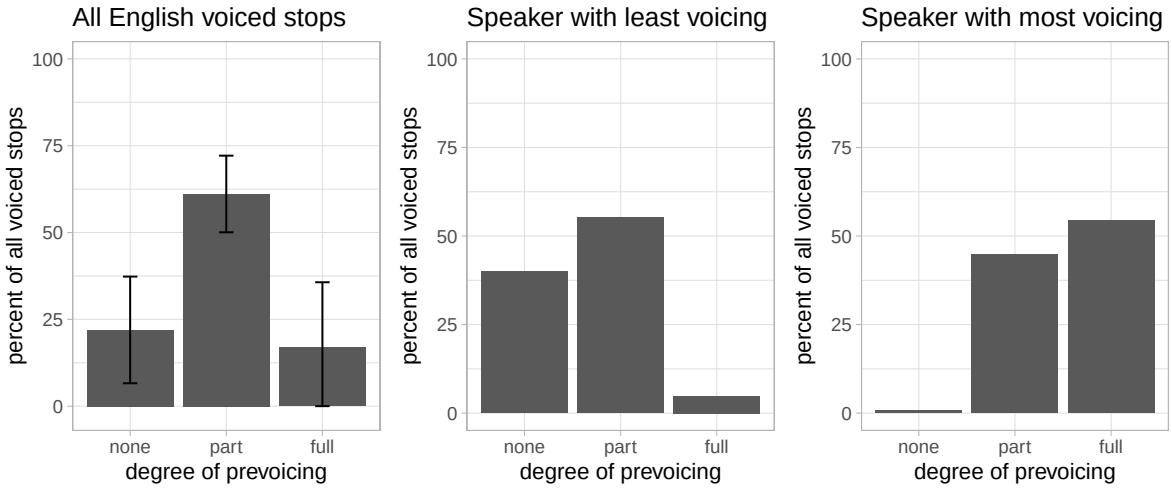


Figure 2.13. Prevoicing across vowel contexts in Hindi phonologically voiced stops

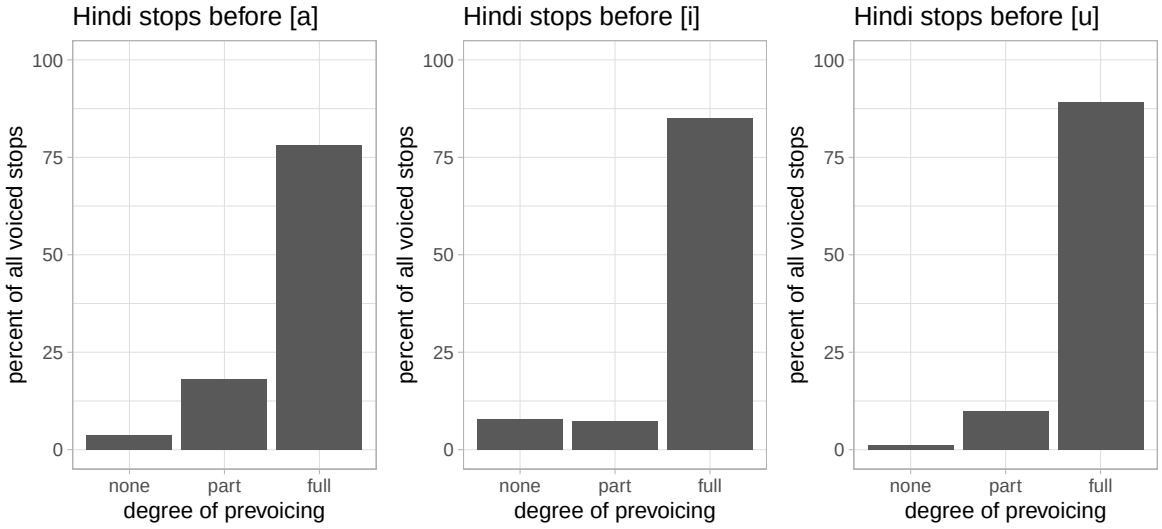
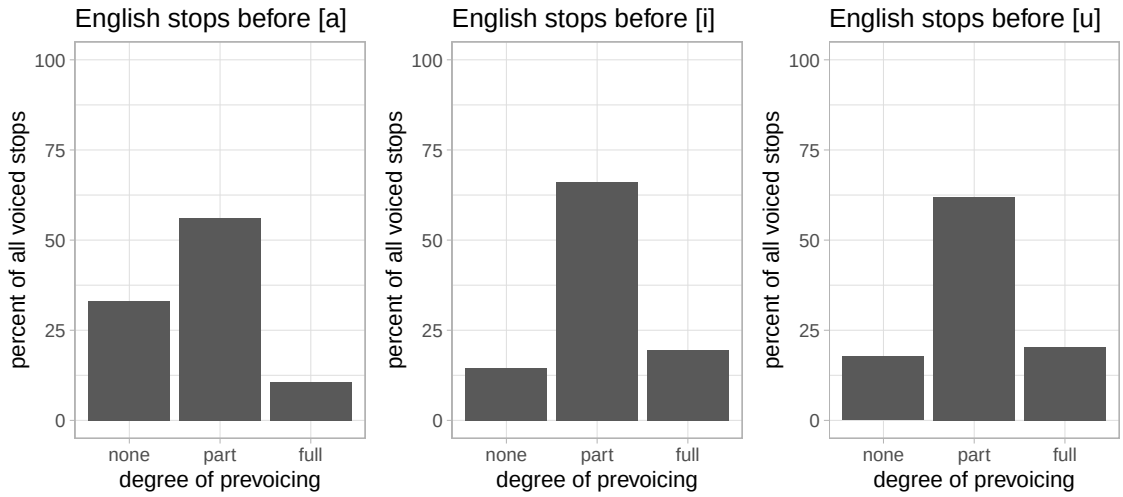


Figure 2.14. Prevoicing across vowel contexts in English phonologically voiced stops



distributions, the Beta Regression assumes Beta distributions, which tend to be more characteristic of proportion data. As evident from the density plots in the previous section, the data here are not normally distributed and would be better approximated with Beta distributions.

Separate regression models were fit for English and Hindi using the following factors as predictors: phonological voicing, place of articulation, speaker, and vowel context (V), experimental block, and closure duration, with random intercepts for word. The following interactions were also included in the full models: place × V, place × speaker, and V × speaker.

Models were fit using the betareg (Cribari-Neto and Zeileis, 2010) and glmmTMB (Brooks et al., 2017) R packages. Best fit models for both languages were determined using variable selection with the Akaike Information Criterion (AIC) (Akaike, 1974). Likelihood ratio tests were performed using the lmtest package (Zeileis and Hothorn, 2002) and stepwise selection was performed using the MASS package (Ripley et al., 2013).

While using fixed effects to model factors like vowel context or word type is typical, speaker effects are often modeled using random effects (Allen et al., 2003; Baayen

et al., 2008). However, there are multiple reasons why a fixed effect for speaker is preferable for this analysis. First, random effects are best suited to factors which have many levels. A standard recommendation is at least five levels (Baayen, 2008; Gorman and Johnson, 2013; Barr, 2013; Coolican, 2017). In this experiment, speaker has six, which is small for accurately estimating variance from a random effect.

In addition, the main question here is how the sources of variance differ in the two languages. Including speaker as a fixed effect allows for the quantitative measurement (via the R squared value) of how much variation is accounted for by speaker relative to the other factors. We are less concerned with which main effects are significant (as is typical when using speaker as a random effect) and instead more concerned with how much variance is accounted for by each factor and how the best fit models differ between languages. If speaker is included as a random effect, the pattern of results is consistent: there is more voicing variation in English relative to Hindi. This alternative analysis is provided in Appendix A.3.

Regression models were fit for both languages using the full effect structure described above (full output in the appendix in Tables A.1-A.2). There is a significant effect of phonological voicing in both languages. In English, there is a significant effect of the high vowel /i/ (indicating more voicing relative to /a/), but neither of the vowel effects could be considered even marginally significant in Hindi. All English speakers show significant or marginally significant speaker effects, while there is only one speaker with a marginally significant effect in Hindi. In addition, there are several significant interactions between speaker, vowel context, and place of articulation in English, while none of the interactions reach significance in Hindi.

The differences between the full models for the two languages result in different best fit models using the AIC criterion for model selection. The best fit model for the Hindi data (given in Table (2.10)) includes only two of the predictors from the full model: phonological voicing and closure duration. With only these two factors, this

Table 2.10. Effect table for best fit model in Hindi. Beta regression with logit link. Dependent variable: closure voicing. Call: voicing percent ~ phonological voicing + closure duration.

Effects	Estimate (se)	z	p
(Intercept)	3.69(0.09)	43.23	$< 0.001^{***}$
Voicing	-4.92(0.09)	-55.58	$< 0.001^{***}$
Closure duration	-2.26(0.45)	-5.05	$< 0.001^{***}$
Pseudo R^2: 0.78			

Table 2.11. Model comparison: Likelihood ratio test of Hindi restricted model vs. full model

Model 1: full model (voicing percent ~ voicing + V × speaker + place × V + place:speaker + closure duration + block + (1 | word))
Model 2: best-fit model (voicing percent ~ phonological voicing + closure duration)

Model	#Df	Log Likelihood	Change in Df	ChiSq	p
1	41	6239.60			
2	4	6221.10	-37	36.82	0.48

model accounts for 78% of the voicing variation in the data. Vowel context, speaker, block, their interactions, or the random effect of word do not significantly improve the model fit. A likelihood ratio test comparing the full model to the best fit model verifies this (given in Table 2.11). The non-significant Chi Square value indicates that there is no significant change in log likelihood when the full model is reduced to the best-fit model.[10]

The best fit model for the English data includes the same predictors as the best fit model in Hindi (voicing and closure duration) as well as vowel context, speaker, the V × speaker interaction, the place × V interaction, the place × speaker interaction, and experimental block. The model is given in Table 2.12. This model accounts for 40% of the overall closure voicing variation in the English data. The only factor from

[10]This does not fail to reach significance simply because of the large change in degrees of freedom. A comparison of the two models using the English data instead of the Hindi data results in a Chi Square value of 721.3^{***}.

the full model which is not included in the best-fit model is the random effect of word. However, including that effect does significantly improve model fit, as is seen in the significant Chi Square value in a likelihood ratio test comparing the English best fit model to the English full model (Table 2.13).

The differences in the best fit models between the two languages show how the extent and sources of voicing variation differ. In accounting for overall variation in the data, the best fit Hindi model (with only voicing and closure duration as predictors) accounts for 78% of the variation while the best fit English model only accounts for 40% of the overall voicing variation. The variance accounted for by individual factors also differs between the two languages. In the graphs in Fig. 2.15, I show the proportion of total variance accounted for by each individual factor in both languages. In the Hindi model, 77.82% of the overall voicing variation is accounted for by phonological voicing. Hardly any additional variance is accounted for by any of the other factors. In the English model, only 13.83% of the voicing variation is accounted for by phonological voicing while around 21.70% of the overall variation is accounted for by speaker. The other factors each account for less than 1% of the overall variation.

Although the full model for English still only accounts for 40% of the overall variance, this does not necessarily indicate that the remaining 60% percent of the variance is due to random variation. It could be the case that this variation is also structured by additional factors which are not analyzed in these models. What can be concluded from these models is that the factors analyzed here account for less of the overall variance in the English data relative to the Hindi data. In accounting for variance, the strongest predictor of amount of closure voicing in Hindi is phonological voicing whereas the strongest predictor of voicing in English is individual speaker.

Table 2.12. Main effect table for best fit model in English. Beta regression with logit link. Dependent variable: closure voicing. Call: voicing + V × speaker + place × V + place:speaker + closure duration + block. Intercept is speaker e02 voiced coronal /a/ context block 1.

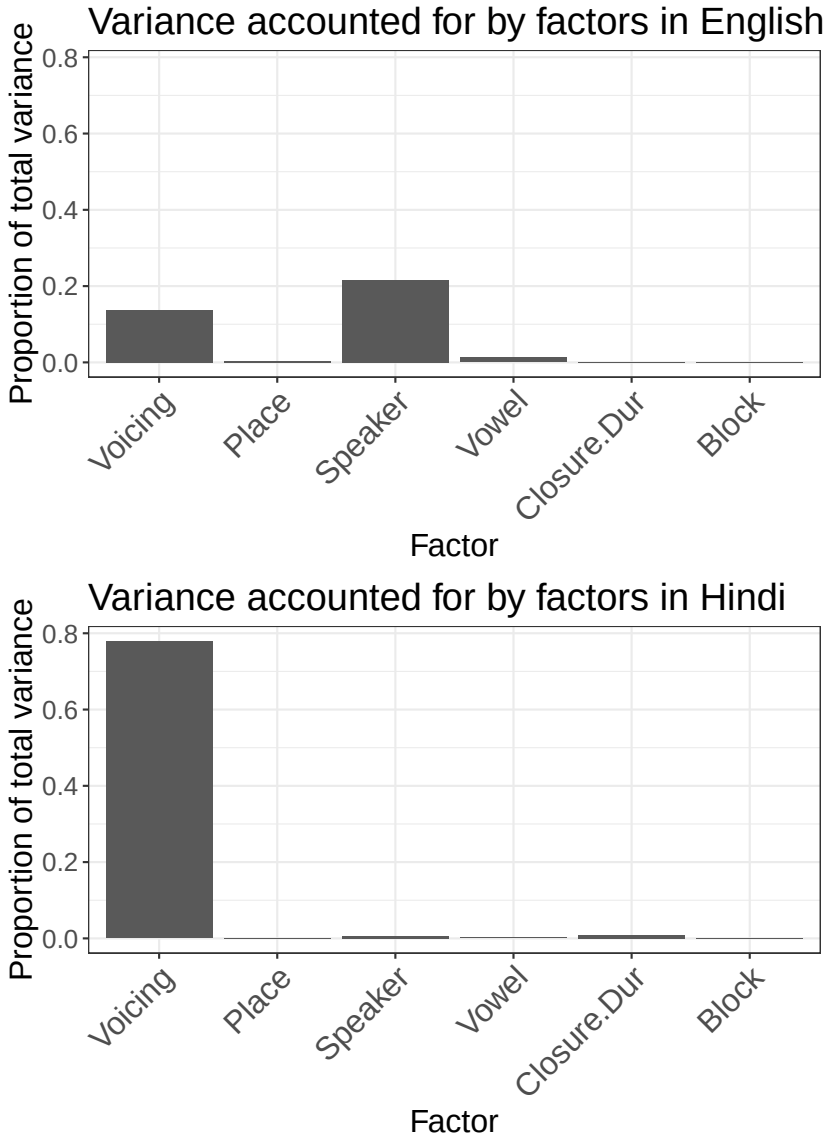
Effects	Estimate (se)	z	p
(Intercept)	2.41 (0.23)	10.34	$< 0.001^{***}$
voicing-voiceless	-1.54 (0.07)	-23.28	$< 0.001^{***}$
V-/i/	0.66 (0.20)	3.24	0.001^{**}
V-/u/	-0.40 (0.21)	-1.86	0.063
speaker-e03	0.55 (0.23)	2.40	0.016^{*}
speaker-e04	-0.66 (0.23)	-2.86	0.004^{**}
speaker-e06	-1.04 (0.23)	-4.56	$< 0.001^{***}$
speaker-e07	-1.94 (0.23)	-8.37	$< 0.001^{***}$
speaker-e09	-1.90 (0.23)	-8.30	$< 0.001^{***}$
place-labial	-0.48 (0.20)	-2.35	0.019^{*}
place-velar	-0.21 (0.20)	-1.04	0.299
closure duration	-9.99 (1.02)	-9.77	$< 0.001^{***}$
block 2	0.05 (0.08)	0.62	0.535
block 3	-0.05 (0.08)	-0.64	0.525
block 4	0.05 (0.08)	0.61	0.541
V-/i/:speaker-e03	-0.44 (0.25)	-1.79	0.073
V-/u/:speaker-e03	0.83 (0.25)	3.31	$< 0.001^{***}$
V-/i/:speaker-e04	-0.07 (0.25)	-0.27	0.788
V-/u/:speaker-e04	1.37 (0.26)	5.30	$< 0.001^{***}$
V-/i/:speaker-e06	-0.55 (0.25)	-2.20	0.028^{*}
V-/u/:speaker-e06	0.50 (0.26)	1.94	0.053
V-/i/:speaker-e07	0.07 (0.25)	0.29	0.774
V-/u/:speaker-e07	1.00 (0.26)	3.89	$< 0.001^{***}$
V-/i/:speaker-e09	-0.11 (0.25)	-0.45	0.656
V-/u/:speaker-e09	0.84 (0.26)	3.28	0.001^{**}
V-/i/:place-labial	-0.08 (0.17)	-0.49	0.622
V-/u/:place-labial	0.08 (0.19)	0.41	0.682
V-/i/:place-velar	0.24 (0.18)	1.35	0.177
V-/u/:place-velar	0.24 (0.18)	1.34	0.182
speaker-e03:place-labial	0.50 (0.25)	2.01	0.045^{*}
speaker-e04:place-labial	0.11 (0.26)	0.44	0.660
speaker-e06:place-labial	0.37 (0.25)	1.45	0.146
speaker-e07:place-labial	0.70 (0.25)	2.77	0.006^{**}
speaker-e09:place-labial	0.51 (0.25)	2.01	0.044
speaker-e03:place-velar	-0.24 (0.25)	-0.95	0.342
speaker-e04:place-velar	-0.68 (0.26)	-2.64	0.008^{**}
speaker-e06:place-velar	-0.50 (0.26)	-1.94	0.053^{*}
speaker-e07:place-velar	0.02 (0.25)	0.09	0.931
speaker-e09:place-velar	-0.11 (0.25)	-0.42	0.674
Pseudo R^2: 0.40			

Table 2.13. Model comparison: Likelihood ratio test of English restricted model vs. full model

Model 1: full model (voicing percent ∼ voicing + V × speaker + place × V + place:speaker + closure duration + block + (1 — word))
Model 2: best-fit model (voicing percent ∼ voicing + V × speaker + place × V + place:speaker + closure duration + block)

Model	#Df	Log Likelihood	Change in Df	ChiSq	p
1	41	1958.40			
2	40	1954.0	-1	8.87	0.003**

Figure 2.15. Proportion of total variance accounted for in regression models



2.5.4 Interim discussion: Voicing

This section has provided results for closure voicing in Hindi and English phonologically voiced stops. I have shown that there is more group-level within-speaker within-category variation in closure voicing in English relative to Hindi. I have also shown that the voicing variation is structured differently in the two languages.

The pattern of voicing in Hindi is consistent across speakers and vowel contexts. The English speakers vary more in closure voicing both within- and between-speakers. Beta regression models for each language show that while underlying phonological voicing accounts for almost 80% of the total voicing variation in Hindi, it accounts for only 13% of the total voicing variation in English. Speaker, vowel context, and their interactions significantly contribute to the English model, showing that the additional variation in English is structured according to these phonologically non-contrastive factors. However, these factors together still account for only 40% of the overall voicing variation in the English data. This suggests that there is either more random variation in English voicing relative to Hindi, or there are additional factors that structure the English variation which are not considered in these models.

This section discusses the voicing results in light of the literature on prevoicing in English stops, reviews implications for laryngeal realism and English featural analyses, and discusses the potential articulatory explanation for variation across vowel contexts.

2.5.5 Prevoicing in English stops

The prevoicing variation in English observed here is in line with recent work on American English documenting prevoicing. Most studies of prevoicing have concentrated on Southern varieties, sometimes reporting prevoicing with higher incidence among male and African-American speakers (Jacewicz et al., 2009; Elston et al., 2016; Herd et al., 2016; Hunnicutt and Morris, 2016). However, none of the speakers in this

study were speakers of a Southern variety.[11] All speakers were female so gender effects could not be tested with the data obtained in this experiment. This suggests that prevoicing in English may be more widespread than previously documented. The results here differ from some (though not all) of the previous studies on English prevoicing in that the stops were elicited intervocalically and not utterance initially. Further work will need to be done with non-Southern populations to see if the prevoicing patterns observed here are also present on utterance-initial stops.

It is possible that the degree of prevoicing observed here is the result of hyperarticulation in a lab setting. However, multiple studies of clear/careful speech have shown that English speakers do not generally prevoice more in these contexts, but instead produce more salient release bursts (Keating, 1984; Picheny et al., 1986; Ohala, 1995; Hazan and Simpson, 2000). While it remains a possibility that the speaker-specific preferences we observed in prevoicing are restricted to the lab context, the existing literature documenting English prevoicing suggests these findings are typical for American English speakers.

It might be the case that the between-speaker variation is due to speaker-specific preference for different hyperarticulation strategies. Speakers which produced mostly prevoiced stops would be using voicing as a way of hyperarticulating voiced stops, and speakers who produced little voicing would be using other strategies (more salient release bursts, increase in lag time difference, etc.). If these differences result from hyperarticulation we may also expect block effects, with hyperarticulation decreasing throughout the experiment. I did not observe any block effects (when included in the regression models for Hindi and English, block was a non-significant factor in both).

We would also expect to see other evidence of hyperarticulation such as extended lag time on voiceless stops. However, the speakers who produced the most prevoicing

[11] We assume this is the case based on participant answers to a demographic questionnaire. None listed any southern states as places where they or their parents learned to speak English.

did not also produce the longest lag times. In addition, these speakers showed a general preference for prevoicing across all stops, even phonologically voiceless stops. If the speakers who typically exhibit closure voicing during phonologically voiced stops were doing so to hyperarticulate those stops, we would not expect the same speakers to produce voicing during the closure for phonologically voiceless stops. This seems to indicate that these speakers have a more general preference for voicing which (while it might be enhanced in a hyperarticulation context) cannot be solely attributed to hyperarticulation in the lab context.

2.5.5.1 Laryngeal realism and English featural analyses

The English prevoicing results have potential implications for considering laryngeal realism (e.g. Honeybone 2005, Beckman et al. 2013; cf. Cyran 2014) in the featural representation of English stops. Under a laryngeal realist hypothesis, the feature system in phonology should represent the phonetic reality of production. English is frequently analyzed (by laryngeal realists) as a language which does not use the feature [voice], but instead [spread glottis]. This [spread glottis] feature reflects the difference between voiceless short lag stops and voiceless long lag stops. Hunnicutt & Morris (2016) offer a potential laryngeal realist phonological analysis of English prevoicing based on data from Southern speakers.

The fact that English speakers use prevoicing on stops (at least sometimes) is compatible with either a laryngeal realist or relativist analysis and these data do not provide particular counterexamples to previous analyses of English in either framework. However, the individual differences observed in the English prevoicing patterns here potentially suggest different feature specifications on the individual level. For example, the speaker who prevoiced almost all phonologically voiced stops could be described as utilizing both [voice] and [spread glottis] representations simultaneously (as in Hunnicutt & Morris's analysis), while the speaker who prevoiced almost no

phonologically voiced stops could be described as utilizing only the [spread glottis feature].

2.5.5.2 Variation across vowel contexts

As in Smith and Westbury (1975), I observed more prevoicing before high vowels in English. Smith and Westbury (1975) proposed a possible articulatory explanation: moving the tongue root to produce a high vowel puts additional tension on the vocal folds, making it easier to sustain voicing through the closure. However, the Hindi speakers are consistent in voicing across vowel contexts and do not prevoice less in front of low vowels relative to high vowels. The lack of even a small effect of this kind in Hindi suggests two explanations. (1) It could be that the pattern observed in English does not actually have a physiological basis and is a learned non-contrastive pattern or (2) the Hindi speakers are able to overcome the physiological challenges to maintain the contrasts of their language.

2.6 Discussion

2.6.1 Lindblom (1986) and Dispersion Theory

Lindblom's (1986) hypothesis "that phonetic values of vowel phonemes should exhibit less variation in small systems than in large systems" is often assumed to be true, despite scant and conflicting evidence from the literature on vowels (see Ch. 1 for an overview). I argue that we cannot generalize this intuition outside of the F1/F2 space assumed for vowels without a more explicit operationalization of the hypothesis.

My results show that it is not the case that phonetic values in larger "systems" are always less variable. In the experiment here, Hindi speakers showed just as much variation as English speakers in voiceless lag time, despite having twice the number of stop phonemes. Instead, the hypothesis should be defined over single phonetic

dimensions as phonological contrasts are not unidimensional in phonetic space. We expect Hindi speakers to exhibit less variation than English speakers but only along the particular phonetic dimensions which realize additional contrasts.

2.6.2 Links to perception

Lindblom's original hypothesis (and work in DT more generally) assumes that production is optimized for ease of perception through sufficient dispersion of phonological categories in acoustic space. DT hypotheses propose that category overlap is avoided through less within-category dispersion and increased between-category dispersion of mean values in the phonetic space. Modeling work on cue-weighting in perception has shown that weighting cues based on how reliably they distinguish phonological contrasts mirrors the cue-weighting patterns observed in perceptual data (Toscano and McMurray, 2010). The model employed by Toscano and McMurray (2010: 438) estimates the reliability of a phonetic dimesion with a ratio of mean values to within-category variances. This type of model is supported by empirical work on the relationship between within-category variability and cue-weighting in perception. Clayards et al. (2008) showed that perceptual uncertainty increases with within-category phonetic variability.

The results of this experiment provide empirical support from production for the inclusion of within-category variance in cue-weighting models. A prediction that arises from the reliability definition in Toscano and McMurray (2010) is that strength of cue and relative amount of within-category variation should be inversely correlated. The perception literature has shown lag time to be a higher weighted cue relative to prevoicing for the phonological voicing constast in English (among others, Lisker and Abramson (1964); Shultz et al. (2012)). In the results here, English speakers exhibited more variation on the prevoicing dimension (a secondary perceptual cue to

the stop contrast in English) relative to Hindi speakers, for whom prevoicing provides a primary perceptual cue.

2.7 Conclusion

This experiment compared acoustics of stop production in Hindi and English. Hindi and English speakers produced similar amounts of within-category variation in voiceless lag time, but English speakers produced more variation in closure voicing. This is in accordance with Contrast-Dependent Variation, my proposed revision of Lindblom's (1968) hypothesis: there should be less variation along a phonetic dimension in languages that realize more phonological contrasts along that dimension relative to language that realize fewer contrasts on that dimensions (Table 2.5).

While it is well-established that production is variable in every language, these results show that patterns and sources of variation are language-specific and relative differences can be predicted. Despite physiological constraints, speakers can constrain variation to preserve phonological contrast, as in Hindi. Speakers allow variation along dimensions which do not threaten phonological contrast and this variation is can be structured according to non-contrastive patterns.

CHAPTER 3

BETWEEN-LANGUAGE CASE STUDY: SIBILANT FRICATIVES IN POLISH AND FRENCH

3.1 Introduction

In Chapter 2, I examined differences in extent of within-category phonetic variation in stops between Hindi and English. The main finding was that speakers of Hindi did show less variation than speakers of English, but only along the dimension of closure voicing. This chapter extends the Contrast-Dependent Variation hypothesis to another case study: sibilants in French and Polish. I present the results of a speech production experiment matching the methods of the stop experiment as closely as possible. The sibilant case is similar to the stop case in that one language has more phonemes than the other and Contrast-Dependent Variation makes different predictions from Lindblom (1986).

Under a direct implementation of Lindblom (1986), we might consider the sibilant inventory to be the relevant space for comparison across the two languages. Polish has more sibilant phonemes than French, therefore we would predict generally less within-category variation in Polish. Contrast-Dependent Variation investigates individual phonetic dimensions instead of inventory subsets. In this chapter, I examine variation along two phonetic dimensions, spectral center of gravity and the second formant at vowel onset. Although there are differences in the number of sibilant categories, both languages use the COG and F2 dimensions as primary cues to phonological contrasts. In French, F2 is relevant for vowel contrasts instead of sibilant contrasts. However, as Contrast-Dependent Variation is implemented over phonetic dimensions

Table 3.1. Consonant inventory of French (Fougeron and Smith, 1993)

	Labial	Dental	Postalveolar	Palatal	Velar	Uvular
Stop	p b	t d			k g	
Fricative	f v	**s z**	**ʃ ʒ**		(x)	
Nasal	m	n		ɲ	(ŋ)	
Approximant		l		j	w	ʀ

instead of inventory subsets, there is still no difference in variation expected between the two languages. The null results presented here clarify the implementation of the Contrast-Dependent Variation predictions: the relevant "space" for evaluating the hypothesis must be defined over phonetic dimensions rather than subsets of the phonemic inventory to accurately capture the patterns of within-category variation between-languages.

3.1.1 French background

French is described as having a voicing contrast in sibilants at two places of articulation (Fougeron and Smith, 1993). The consonant inventory of French is shown in Figure 3.1. French /s z/ are typically described as having dental and apical articulation and French /ʃ ʒ/ are typically described as having palato-alveolar or postalveolar articulation, although there is some disagreement about this in the literature. Dart (1998) provides a review of the claims about sibilant place articulation in the literature on French. Dart (1998) also presented results from a comparative study of coronal articulation in English and French. The group patterns for articulation of French /s/ and /ʃ/ were dental and postalveolar repectively. However, French speakers did exhibit individual variation in place and apicality of [s]. In this chapter, I follow Dart (1998) in referring to the two fricatives as dental and postalveolar.

There have been relatively few studies examining perception of French fricatives by adult listeners.[1] McCasland (1983) used a fricative identification task to investigate which cues contribute to fricative discrimination in French. They showed that while both noise intensity and spectral cues were used to discriminate sibilants and non-sibilants, spectral mean/center of gravity (COG) was mostly used to discriminate the two sibilants. This is in line with work on English, which has a similar sibilant system (Jongman et al., 1989; Jongman and Wade, 2007). Such work also shows that F2 of following vowels covaries with sibilant category in English and can be understood as a secondary cue to the sibilant contrast. These differences likely also occur in French. Following McCasland (1983) and more recent work on sibilant assimilation in French (Clayards et al., 2015), I assume COG to be the primary phonetic cue distinguishing the dental sibilant from the postalveolar sibilant in French.

3.1.2 Polish background

Polish has been described as having contrastive coronal fricatives at three places of articulation: alveolar, alveopalatal, and retroflex (Dogil, 1990). The consonant inventory of Polish is shown in Table 3.2. Multiple studies (Nowak, 2006; Bukmaier et al., 2014) report on a small number of speakers who show a spectral center of gravity (COG) contrast between the dental fricative and the other two fricatives in production. Several studies have reported little difference in spectral center of gravity and other spectral measures of the frication noise (Żygis and Hamann, 2003; Bukmaier and Harrington, 2016; Lee-Kim, 2011). Instead, the alveopalatal and retroflex

[1]There have been studies of fricative perception by French-acquiring infants (Cristià et al., 2011) as well as studies investigating the perceptual effects of place assimilation in sibilant sequences (Niebuhr et al., 2008, 2011; Clayards et al., 2015). These are not particularly relevant to the present study as our participants were adults and all of the sibilants elicited were intervocalic.

Table 3.2. Consonant inventory of Polish (Padgett and Żygis, 2007)

	Labial	Alveolar	Alveopalatal	Retroflex	Palatal	Velar	Glottal
Stop	p b	t d				k g	
Fricative	f v	**s z**	**ɕ ʑ**	**ʂ ʐ**		x	h
Affricate		ts dz	tɕ dʑ	tʂ dʐ			
Nasal	m	n			ɲ		
Approximant		l ɹ			j	w	

fricatives have been described as being distinguished by the transition of the second formant (F2) into the following vowel.[2]

When reporting on differences in vowel transitions, some quantify the vowel transition by the F2 difference at two vowel timepoints (Nowak, 2006; Chiu, 2009); some quantify the transition by reporting the F2 value at vowel onset (Halle and Stevens, 1997; Kudela, 1968). Bukmaier and Harrington (2016) analyze onset of vowel transitions and show higher F2 values in the vowels following the alveopalatal, but these values showed some overlap with vowels following the retroflex. They report F2 trajectories between the onset and midpoint of the following vowel and argue that the raised F2 values following the alveopalatal are evidence for a coarticulatory palatalizing influence.

In perception, Nowak (2006) showed that frication noise alone is sufficient to categorize isolated fricatives for native speakers of Polish. However, it is possible the speakers are not interpreting the isolated fricatives as speech and therefore have enhanced discriminability. When the Polish fricatives were placed in a VCV context, removal of the formant transitions into the following vowel made the alveopalatal and retroflex fricatives confusable, indicating the primacy of F2 over COG as a cue to the contrast between /ɕ/ and /s ʂ/.

[2]Żygis and Hamann (2003) also report data from a second speaker who has a three-way COG contrast between the sibilants.

It has been argued that the Polish three-way sibilant contrast is diachronically unstable. These arguments tend to center the retroflex as being particularly unstable and predict that it will merge with either the alveopalatal or the dental Bukmaier et al. (2014); Żygis et al. (2012). Mergers of both types have been reported in some nonstandard dialects of Polish (Żygis et al., 2012; Nowak, 2006; Bukmaier et al., 2014).

Nowak (2006) and Bukmaier et al. (2014) argue that the dental-retroflex merger is more likely. The argument is based on: sensitivity to formant transitions in acquisition and increased variability in retroflex articulation. Under the argument that acquisition drives sound change (e.g. Stampe, 1972; Greenlee and Ohala, 1980; Blevins, 2004), we anticipate the retroflex-alveopalatal contrast to be more stable (as it is distinguished primarily by formant transitions) than the retroflex-dental or alveopalatal-dental contrasts which are distinguished primarily by COG. This is because acquisition data show that children weigh formant transition information higher than COG information in perception (Nittrouer and Studdert-Kennedy, 1987).

The retroflex is also argued to be particularly unstable due to increased articulatory variability relative to the other sibilants. Bukmaier et al. (2014) use EMA to investigate tongue shapes of Polish speakers. They show that articulation of the retroflex fricative varies more with speech rate than the other fricatives. At slower speech rates, the retroflex displayed a sub-laminal production while at faster speech rates the the tongue tip orientation was super-laminal. The other fricatives did not show a comparable difference between speaking rates. This is potentially similar to the retroflex in Mandarin Chinese. Hu (2008) also showed greater articulatory variability for /ʂ/ relative to the other sibilants in Mandarin. I return to the comparison of Polish and Mandarin fricatives in Chapter 4 after the presentation of the Mandarin data.

While both Bukmaier et al. (2014) and Hu (2008) examine only articulatory variability and not acoustic variability, the increased articulatory variability could result in increased acoustic variability. However, the Polish variation was analyzed across two speech rate conditions. In the present study, there is no manipulation of speech rate, so we do not necessarily expect greater variability for the retroflex in the present results.

Although the contrast is argued to be diachronically unstable, many modern dialects of Polish maintain the three way contrast. Boersma and Hamann (2008) models a potential path of diachronic development of the contrast resulting in synchronic stability. They show that when production and perception are modeled with bidirectional phonetic cue constraints, dispersion is emergent without constraints demanding dispersion. The non-dispersed inventories (exaggerated, confusable, or skewed contrasts) are not stable over time and move towards more dispersed configurations in the diachronic simulations.

The phonetic cue constraints in Boersma and Hamann (2008) modulate between the phonological surface form (the output of the phonological grammar) and the realized phonetic form. The cue constraints are similar to faithfulness constraints in standard OT in that they evaluate the relation between the phonological and phonetic forms. The articulatory constraints are similar to markedness constraints in that they only evaluate the phonetic form and not the relation between the two forms. However, articulatory constraints are universally ranked because they reference invariant articulatory difficulties which do not vary cross-linguistically.

In the Boersma and Hamann (2008) model, it is assumed that the learner has correct lexical representations but has not yet acquired pre-lexical phonetic perception. With the use of a lexicon, the learner acquires the correct ranking of phonetic cue constrains for their language using the Gradual Learning Algorithm (Boersma and Hayes, 2001). This model is crucially bidirectional, meaning that the same constraints

are used for perception and production. This is argued to be the main factor which causes emergent dispersion over time.

They use Polish as an example case of a sibilant inventory which their model predicts to have emergent dispersion across the three categories over time. However, this result is not totally analogous to the actual phonetic facts about Polish sibilants. While they accurately model the instability of Medieval Polish, the model does not accurately reflect modern Polish sibilants. The model only considers dispersion across one phonetic dimension, spectral center of gravity. Over time, the sibilants become more dispersed across this single dimension.

This is presented as a desirable result, yet work on modern Polish (summarized above) has shown that the sibilant contrast is instantiated over multiple phonetic dimensions. While some speakers have a three way COG sibilant contrast, many speakers (including the ones in the present study) employ both the COG dimension and the F2 dimension to make the full three way contrast. The use of additional phonetic dimensions to achieve a dispersed inventory is not predicted by the model formulated in Boersma and Hamann (2008), although a reformulation could potentially include multiple dimensions. It is possible that a more complex model incorporating multiple phonetic dimensions would more accurately simulate the actual development of Polish sibilants as contrasted over two phonetic dimensions. The results from Boersma and Hamann (2008) show one way in which dispersion of mean acoustic values could arise over time. Their model does not necessarily make any predictions about extent of within-category variation, which is examined here.

In sum, the three way sibilant contrast in Polish is often argued to be diachronically unstable, yet modeling work shows the potential for stability through emergent acoustic dispersion. However, as summarized above, recent work on modern Polish sibilants indicates that many speakers instantiate the three way sibilant contrast across two dimensions of COG and F2, rather than dispersed along the COG di-

mension. This is corroborated in perception; listeners use F2 as the primary cue for distinguishing the alveopalatal sibilant from the other sibilants and COG as the primary cue for distinguishing the dental sibilant from the other sibilants. In the analysis here, I assume F2 to be the primary cue distinguishing the alveopalatal from the retroflex and alveolar sibilants and COG to be the primary cue distinguishing the dental from the alveopalatal and retroflex sibilants.

3.1.3 Predictions

Unlike previous literature in Dispersion Theory, this hypothesis is evaluated over phonetic dimensions instead of phonological inventories. The relevant "space" or "system" is a single phonetic dimension rather than a subset of the phonemic inventory (see Table 2.5 for a comparison of this hypothesis with Lindblom (1986)). In this chapter, I test the Contrast-Dependent Variation hypothesis over two phonetic dimensions: COG and F2.

In the Hindi and English case from Chapter 2, the results showed more variation in English, but only in voicing. This was not predicted by the direct implementation of Lindblom (1986) where the stop inventory would likely be the relevant space. This result is predicted under Contrast-Dependent Variation, which considers each phonetic dimension as its own space. There is one contrast in both languages that uses lag time as a primary cue, but there are no contrasts in English which use closure voicing as a primary cue.

We can compare this case with the Polish and French sibilants. Under the most direct implementation of Lindblom (1986), we would consider the sibilant inventory to be the relevant space, and we would generally expect more variation in French as French as fewer sibilant phonemes. Under Contrast-Dependent Variation, we only expect more variation along the particular phonetic dimensions that realize fewer contrasts. Both languages employ the COG dimension as a primary cue to a phonological

contrast. Therefore, we do not expect French speakers to exhibit more within-category variation in COG relative to Polish speakers.

Considering the F2 dimension, previous work shows that Polish speakers use F2 at vowel onset as a primary cue to the sibilant contrast (see summary in §3.1). However, French speakers also utilize F2 as the primary cue to a phonological contrast, but in French it is a vowel contrast, not a sibilant contrast.[3] Contrast-Dependent Variation then predicts no difference in F2 variation between the two languages, as both languages use F2 as the primary cue to a phonological contrast. While Lindblom (1986) predicts a difference between these two languages, Contrast-Dependent Variation predicts no differences along these phonetic dimensions. I return to further discussion of the implications for implementation of the Contrast-Dependent Variation hypothesis in §3.5.

3.2 Experimental design

The methods were matched as closely as possible to the methods used in Chapter 2. Adaptation for French and Polish is described here.

3.2.1 Participants

The French speakers were undergraduate and graduate students at the University of Massachusetts Amherst and surrounding colleges. All French speakers acquired French natively in a predominantly French language environment before relocating to the United States. Seven speakers of French were recorded. Exclusion was determined based on the same criteria used in Experiment 1. Two speakers were excluded due to frequent speech errors and one speaker was excluded due to frequent pauses before

[3]See Strange et al. (2007) for acoustic evidence showing the F2 vowel distinctions made by French speakers and Gottfried and Beddor (1988) for perceptual evidence showing importance of F2.

the stimulus (see Chapter 2 for additional details on the exclusion criteria). After these exclusions, data from four speakers were analyzed.

The Polish speakers were all undergraduate students at the University of Massachusetts Amherst. All Polish speakers which were analyzed acquired Polish natively in Poland before relocating to the United States. Six speakers of Polish were recorded. One Polish speaker was excluded due to frequent speech errors, one Polish speaker was excluded due to long pauses, and one Polish speaker was excluded due to non-native speaker status. After these exclusions, data from three speakers were analyzed.

3.2.2 Stimuli

The stimuli were words and non-words where the onset consonant was a sibilant and the following vowel was one of [ɛ a ɔ]. This differs from the vowel contexts in Experiment 1 due to allophony and neutralization in the [i a u] vowel contexts. Voiced and voiceless sibilants were elicited, but only the voiceless sibilants are analyzed here. The stimuli also differ from Chapter 2 in that they were also crossed over number of syllables. Due to the rich inflectional system in Polish, the lexicon does not include many monosyllabic words. Therefore, the stimuli sets were constructed to include mono-, di-, and trisyllabic words in both languages.

The stimuli were also cross-classified according to word frequency. In French, the Lexique online lexicon (New et al., 2001) was used for selecting stimuli. The words were classified into two frequency categories, high and low. These were determined by native speaker intuitions from a research assistant and verified by the data in Lexique. In Polish, we used the Polimorf lexicon (Wolinski et al., 2012) for selecting

stimuli and verified native speaker frequency intuitions with orthographic frequency data.[4]

All the stimuli were recorded in carrier phrases: "Powiedzała X od razu/obecnie" ('She said X right away/now') in Polish and "Il dira X a lui/encore" ('He will say X to him/again') in French. The second half of the carrier phrase was interchangeable and both were used to help retain attention throughout the task. Polish stimuli were crossed according to: sibilant (6 levels) × vowel context (3 levels) ×word status (3 levels: high frequency/low frequency/non-word) × number of syllables (3 levels). Due to gaps in the Polish lexicon, not all factors could be fully crossed[5] and the full stimuli set included 126 stimuli. French stimuli were crossed according to: sibilant (4 levels) × vowel context (3 levels) × word status (3 levels: high frequency/low frequency/non-word) × number of syllables (3 levels) for a total of 108 distinct stimuli.[6]

3.2.3 Recording

Recording was done according to the same procedure used in Chapter 2. After recording the participants completed a word frequency judgment task with the stimuli. This was to ensure that the frequency data matched the intuitions of the participants. The judgment task consisted of filling out a survey indicating degree of familiarity with each word and took between 2-5 minutes to complete.

---

[4]Thanks to Gaja Jarosz for providing this data and offering helpful commentary on the experimental design.

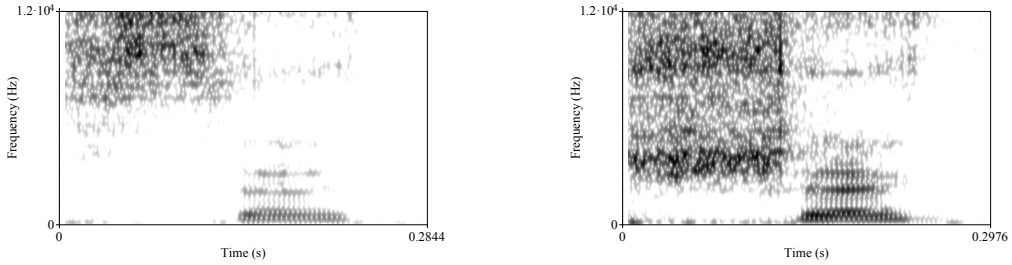
[5]Non-words were used, however, word status is still not fully crossed with all the other factors if there are lexical gaps.

[6]The full stimuli list and other experimental materials are available in the public data archive for this project at https://osf.io/2famr/.

Table 3.3. Example stimuli

Language	C_1	vowel	stimulus (IPA)	type
French	s	ɛ	sɛt	word-hi
French	s	ɛ	sɛʒ	word-low
French	ʃ	ɛ	ʃɛʒ	non-word
French	s	a	saʃ	word-hi
French	ʃ	a	ʃak	word-hi
French	s	a	sad	non-word
French	ʃ	ɔ	ʃɔp	word-hi
French	ʃ	ɔ	ʃɔk	word-low
French	s	ɔ	sɔf	non-word
Polish	ɕ	ɛ	ɕedem	word-hi
Polish	s	ɛ	sɛp	word-low
Polish	s	ɛ	sɛbovali	non-word
Polish	ɕ	a	ɕano	word-low
Polish	ʂ	a	ʂal	word-hi
Polish	s	a	sad	non-word
Polish	ʂ	ɔ	ʂopa	word-hi
Polish	ɕ	ɔ	ɕɔrb	word-low
Polish	s	ɔ	sɔf	non-word
Polish	ɕ	ɔ	ɕorpany	non-word
Polish	ʂ	ɔ	ʂɛf	word-hi

Figure 3.1. Dental and postalveolar sibilants in French. Left panel /sɛ/, right panel /ʃɛ/.

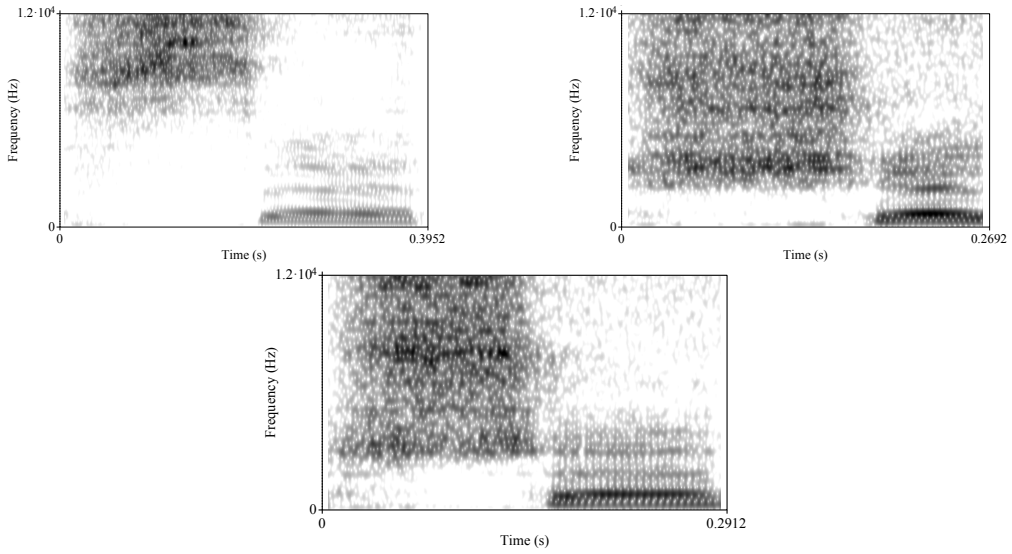


3.3 Analysis

The analysis used similar methods to the analysis in Chapter 2. The recordings from each speaker were first scanned by the author and research assistants for speech errors. The recordings were forced aligned using the Montreal Forced Aligner McAuliffe et al. (2017) by training new acoustic models on the data. A Praat script based on DiCanio (2013) was used to extract spectral moments of the fricatives and formant values of the following vowels. The formants were estimated using the Burg method and extracted at 10 ms intervals throughout the duration of the vowel. Formant excursions greater than 1000 Hz over 10 ms were assumed to be tracking errors and were excluded. This excluded a total of 30 observations across all speakers, sibilants, and vowel contexts.

Example French tokens are given in Figure 3.1. There is a difference in spectral center of gravity in the frication noise, with the dental sibilant exhibiting a much higher COG relative to the postalveolar. Example Polish tokens are given in Figure 3.2. There is also a COG difference here between the dental sibilant and the other two sibilants. The retroflex and the alveopalatal have similar centers of gravity, but the second formant of the vowel is slightly higher following the alveopalatal than the retroflex.

Figure 3.2. Dental, alveopalatal, and retroflex sibilants in Polish. Left panel /sɛ/, right panel /ɕɛ/, bottom panel /ʂɛ/.



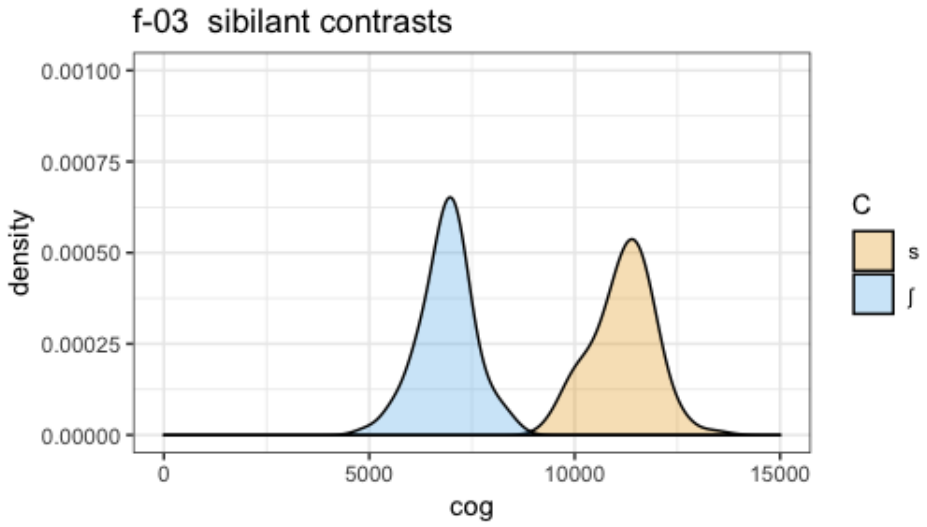
3.4 Results

3.4.1 Center of gravity (COG)

The COG results corroborated the findings in the previous literature. In French, speakers showed a COG distinction between both sibilants. Results from a representative French speaker are shown in Figure 3.3 (results from all speakers are given in Appendix B.1). The figure contains data from one speaker and shows a density plot with two distinct distributions representing the two sibilants categories of French. This speaker has two distinct COG categories for /s/ and /ʃ/ with little overlap between categories, a pattern observed in all speakers of French in this study. Speakers differed in mean values of the sibilant categories but all showed the same pattern of two distinct categories along the COG dimension.

In Polish, speakers showed a COG distinction between the dental sibilant and the other two sibilants. Results from a representative Polish speaker are shown in Figure 3.4 (results from all speakers given in Appendix B.1). The figure is also a density plot showing three distributions representing the three sibilant categories in Polish. This speaker has a two-way contrast between the dental sibilant and the other two

Figure 3.3. Voiceless sibilant COG contrast for a representative French speaker



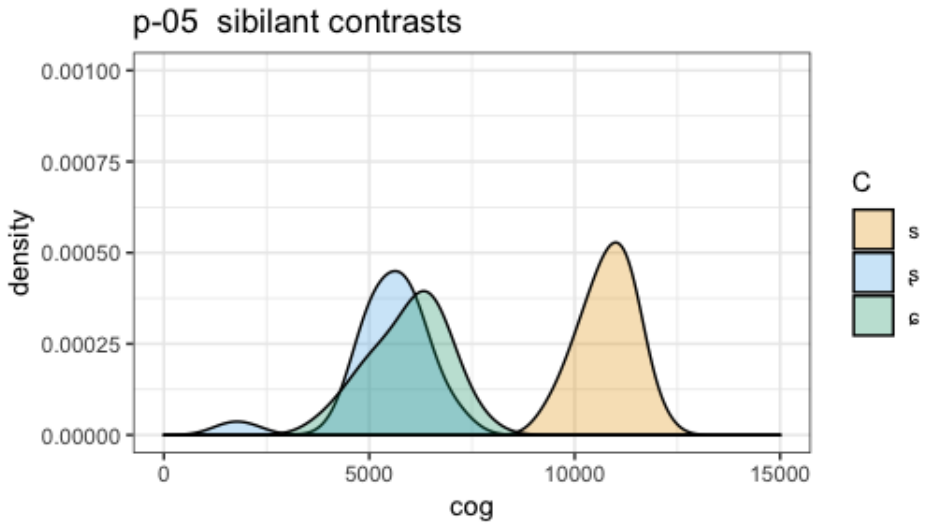
sibilants, with a large degree of overlap in COG between /ʂ/ and /ɕ/. All Polish speakers recorded here displayed a similar pattern with a distinct COG distribution for /s/ relative to the other two sibilants and more COG overlap between the /ʂ/ and /ɕ/. Speakers differed in mean values, variance within categories, and amount of overlap between the retroflex and alveopalatal.

3.4.2 Second formant of the following vowel (F2)

3.4.2.1 French

The French speakers did not produce noticeable differences in vowel formant trajectories following the two sibilants. This is in accordance with the previous literature describing the sibilants as primarily distinguished by COG. Data from a representative French speaker are shown in Figure 3.5 (data from all speakers given in Appendix B.1). The figures show F2 values of the following vowel taken at 10ms intervals. In each vowel context, there is no clear distinction between the F2 trajectories following /s/ and the F2 trajectories following /ʃ/, which was the case for all speakers of French. While mean F2 values differed across speakers, all speakers displayed a similar pattern of overlapping trajectories following the dental and postalveolar sibilant.

Figure 3.4. Voiceless sibilant COG contrast for a representative Polish speaker



3.4.2.2 Polish

In accordance with the previous literature, Polish speakers produced differences in F2 following the alveopalatal sibilant relative to the other sibilants (see §3.1 for a review). However, the speakers here differ from the previous literature in that the formant distinctions are not consistently in the slope of the formant transition; rather, the onset F2 value is higher following the alveopalatal sibilant and these differences persist throughout the duration of the vowel.

Vowel formant trajectories in Polish for tokens where [ɛ] follows the sibilant are shown in Figure 3.6 for two speakers. Speaker 3 is a male speaker and Speaker 5 is a female speaker (data from all Polish speakers given in Appendix B.1). These two speakers are representative of the group; all Polish speakers show higher F2 values following the alveopalatal relative to the other sibilants and the difference continues throughout the duration of [ɛ]. The speakers differed in mean formant values and within-category variance (which can be seen in the two speakers shown here), but all speakers showed consistently higher F2 values following the alveopalatal sibilant relative to the other sibilant in all vowel contexts.

Figure 3.5. Formant trajectories from a representative French speaker. [ɔ] context (top panel), [a] context (middle panel), [ɛ] context (bottom panel)

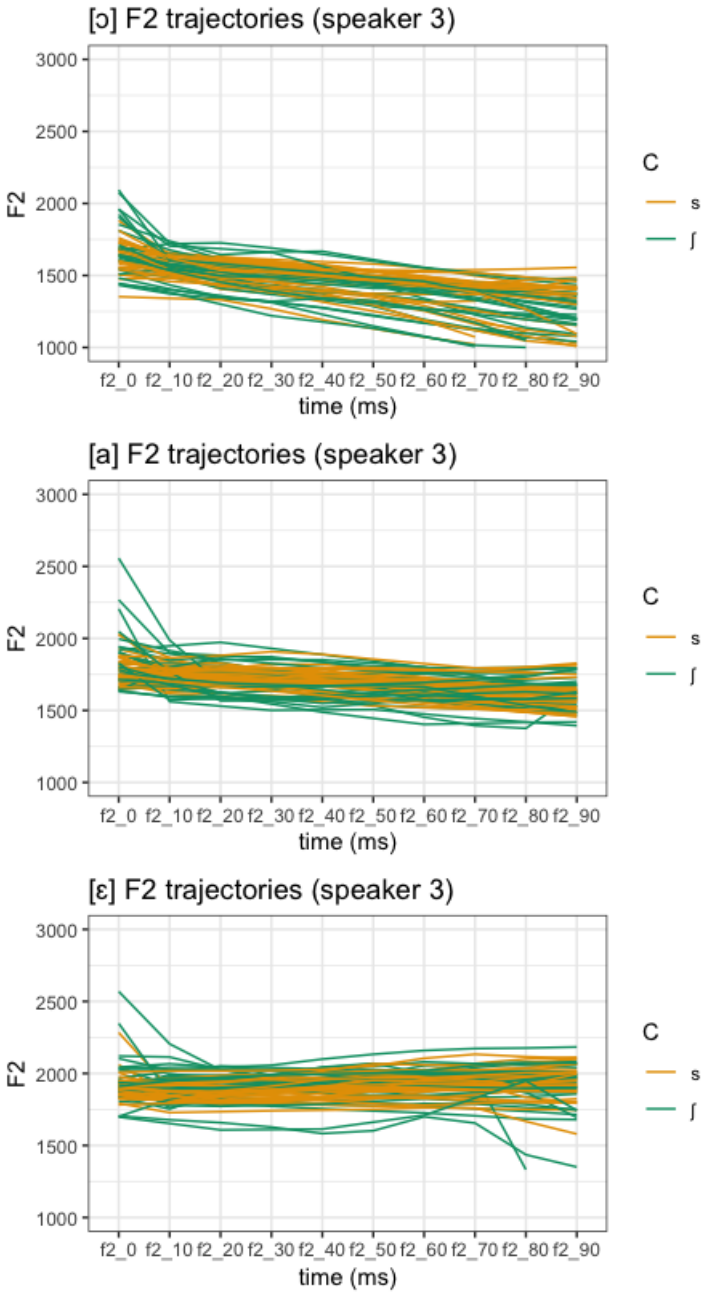
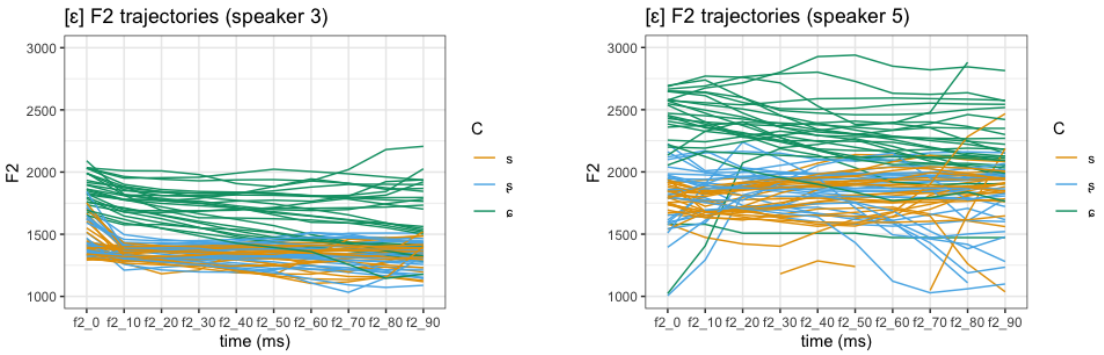


Figure 3.6. F2 trajectories for [ɛ] following /ɕ/ and /ʂ/ in Polish.



There was more between-speaker variation in the other vowel contexts /ɔ a/. Some speakers showed rising transitions into /ɕ/ with differences that did not continue until the end of the vowel (much like what has been described in the previous literature), while other speakers showed flat transitions where the differences were sustained throughout the entire vowel duration.

For the hypothesis under investigation here, we are interested in which dimension is the primary correlate of the contrast between the alveopalatal sibilant and the other two sibilants. Although the exact shape of the formant transitions differs from what has been found in the previous literature, these findings are in accordance with the perceptual finding that the primary cue to the alveopalatal-retroflex contrast is not in the spectral cues of the sibilant, but in the formants of the following vowel. In the analyses that follow, I take F2 at vowel onset of the following vowel to be the phonetic dimension of interest.

3.4.3 Comparative results

Under Lindblom's (1986) hypothesis, we would expect more F2 variation in French than in Polish if we consider the sibilant inventory to be the relevant system as Polish has more sibilant phonemes than French. If we instead consider phonetic dimensions as the relevant systems (as proposed here with Contrast-Dependent Variation), we

do not expect any difference as both languages use the COG and F2 dimensions as primary cues to phonological contrasts.

Following the same method as in Chapter 2 §2.4.2, I use a linear mixed effects regression model where within-category within-speaker F2 variation is the dependent variable. F2 values at 10ms into the following vowel were used for the variance calculations. Under the hypothesis that predicts more variation in French, we expect a significant main effect of language. If a difference in variation is only observed for one of the sibilants, we could expect a significant effect of the language × sibilant interaction.

For the purposes of this analysis, the dental sibilants in both languages are considered to be analogous and are given the same category label, /s/. The postalveolar sibilant could be considered analogous to either the alveopalatal or the retroflex sibilant in Polish. Here, I consider the French postalveolar to be comparable to the Polish retroflex for the purposes of comparing within-category variance. The results do not meaningfully change if the French postalveolar is instead compared to the Polish alveopalatal.[7] The results of the model are given in Table 3.4.

There is no main effect of language in the model. There is a significant main effect of the alveopalatal sibilant, indicating that there is more within-category variation for the alveopalatal sibilant relative to the intercept /s/ (although there is no French analogue so this does not reflect an effect of language). There is also a significant main effect for the vowel /ɛ/, which indicates significantly less variation in /ɛ/ relative to the other vowels. Although there is no significant main effect of language there is a significant effect of the interaction between language and the retroflex/postalveolar categories. This indicates that there is a difference in within-category F2 variation

[7]It could be argued that the only phones which are comparable across the two languages are the dental sibilants. There is no significant difference in extent of within-category variation between the dental sibilant in Polish and the dental sibilant in French.

Table 3.4. Fixed effect table for mixed effects linear regression in Polish and French. Dependent variable: within-category within-speaker F2 variation. Predictors: language, C (sibilant category), V (vowel context), C×language. Model intercept is French /sa/.

Effects	Estimate (se)	t	p
(Intercept)	8.66(2.75)	3.15	0.016*
language-Polish	0.50(4.06)	0.12	0.910
C-/ɕ/	5.78(1.66)	3.49	0.001**
C-/ʃ ʂ/	0.39(1.43)	0.27	0.786
V-/ɛ/	−3.30(1.21)	-2.74	0.009**
V-/ɔ/	0.93(1.21)	0.77	0.45
language-Polish × C-/ʃ ʂ/	5.16(2.19)	2.36	0.024*

between the retroflex sibilant in Polish and the postalveolar sibilant in French, but no difference in variation between the dental sibilants in the two languages. However, this effect is in opposite direction from what was predicted. The hypothesis predicts less variation in Polish and the model shows significantly more variation in Polish /ʂ/ relative to French /ʃ/.

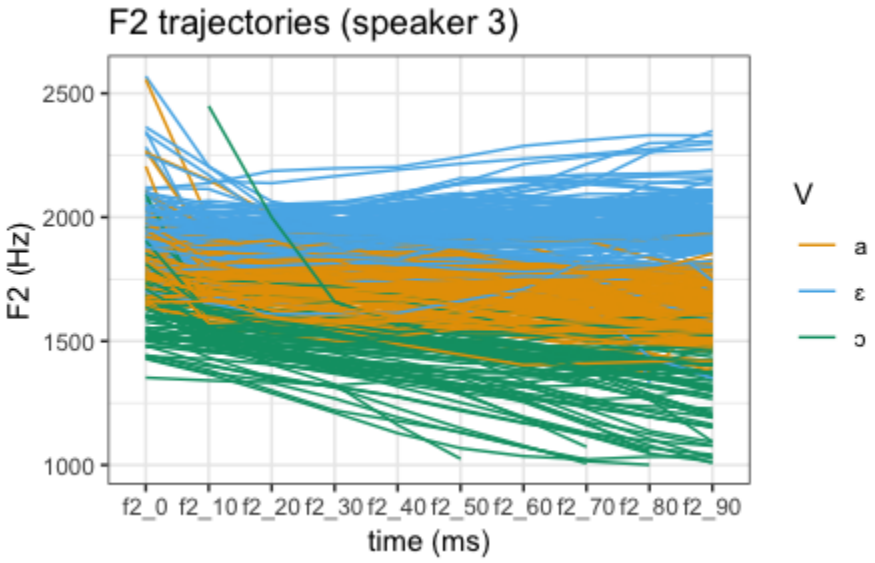
3.5 Discussion

3.5.1 Clarifying hypothesis implementation

Under the Contrast-Dependent Variation hypothesis proposed here, we do not expect a difference in amount of within-category variation in COG or F2 between these two languages. This is because both phonetic dimensions are employed as primary cues to phonological contrasts in the context elicited in the experiment in both languages. In order to determine whether a given dimension is a cue to a phonological contrast in production, we examine the values along that phonetic dimension to determine whether they are predictive of phonological category membership. In other words, we calculate the relative cue weight of that dimension.[8]

[8]This could be done using a variety of statistical methods. Linear discriminant analysis has been used in the phonetics literature to quantify strength of a phonetic dimension as predictive of

Figure 3.7. F2 trajectories across elicited vowel contexts for a representative French speaker.



We know from previous literature (see §3.1 for a review) that both languages use the COG dimension as a primary cue to sibilant contrasts: /s/ vs. /ʂ ɕ/ in Polish and /s/ vs. /ʃ/ in French. F2 is additionally a cue to sibilant place of articulation in Polish and distinguishes between /ɕ/ vs. /s ʂ/. These findings are replicated in the data here. Figures 3.3-3.6 show sibilant category separation along the COG dimension in both languages and sibilant category separation along the F2 dimension in Polish.

While French does not use F2 to distinguish sibilants, F2 is used as a cue to vowel contrasts. The figure in 3.7 shows F2 trajectories in three vowel contexts from a representative French speaker. While there is some overlap between phonological categories, this speaker produces higher F2 values for /a/ relative to /ɔ/ and higher F2 values for /ɛ/ relative to /a/. All French speakers show this pattern, which is expected based on the previous literature documenting F2 as a primary cue to vowel backness (see §3.1 for further discussion of previous literature).

---

phonological category membership, or cue weight in production (e.g. Shultz et al., 2012). These methods are discussed further in Chapter 4, where I employ LDA to determine cue weight for the purposes of testing predictions of Contrast-Dependent Variation in Mandarin sibilants.

I argue that the relevant "space" for considering the predictions of Contrast-Dependent Variation is a phonetic dimension. In this chapter, the phonetic dimensions under investigation are spectral center of gravity (time-normalized over the middle 80% of the fricative) and the second formant of the following vowel measured 10ms after vowel onset. While F2 is often assumed to be a single phonetic dimension, there are many ways of extracting information about F2. Following Yu (2017), I consider a formant value at a temporally defined point of extraction to be a single phonetic dimension belonging to a larger "family of parameters" (p.126).

When the F2 dimension is defined by F2 at a particular timepoint, it is not necessary to specify an additional relevant unit in which strength of cue should be considered (e.g. word, diphone, etc.). Rather, cue primacy for the purposes of evaluating Contrast-Dependent Variation is simply quantified by cue weight at the point of measurement. Because the dimension under consideration here is F2 at 10ms into the vowel, Contrast-Dependent Variation predicts no difference between Polish and French as F2 (at that point) is predictive of phonological category distinctions in both languages.

This result therefore only refers to the particular F2 dimension analyzed. It is possible that other dimensions within the F2 family (F2 extracted at other timepoints, or derived measurements such as transition slope, etc.) would show a difference in within-category variation between these two languages. Whether (and in what direction) we expect differences under Contrast-Dependent Variation would crucially depend on the cue weights of the particular F2 dimension under consideration.

Contrast-Dependent Variation may predict a difference between the two languages if the sibilants were instead in a word-final or pre-consonantal position. In that case, French speakers would still use COG as a primary cue to the sibilant contrasts and Polish speakers may employ COG as a primary cue distinguishing all three sibilant contrasts (potentially suggested by the findings in Nowak, 2006). If the Polish speak-

ers do in fact employ COG as the primary cue to all three sibilant contrasts in this context, Contrast-Dependent Variation would predict less within-category COG variation in Polish relative to French.[9]

This case study provides another example of how defining the relevant "system" according to phonetic dimensions (as proposed here) makes crucially different predictions from the original formuation of Lindblom (1986). Under the most direct implementation of that hypothesis, we might understand the relevant system to be the sibilant inventory. Therefore, we would expect generally more variation in French relative to Polish because French has fewer sibilant phonemes. Even if the the use of multiple phonetic dimensions is acknowledged, as long as the relevant space is defined with inventory subsets, we would still predict a difference in amount of F2 variation between French and Polish because Polish has more sibilant phonemes.

Because Contrast-Dependent Variation is formulated over phonetic dimensions instead of phonological sub-inventories, it captures the fact that the F2 dimension examined here is used as a primary cue to vowel contrasts in addition to sibilant contrasts. Therefore, it does not predict a difference in within-category variation between the two languages. The fact that F2 is used as a primary cue to vowel contrasts is not captured by a direct implementation of Lindblom (1986) which would not necessarily consider vowel contrasts when comparing within-category variation of sibilant inventories.

As the main result in this chapter is null (no difference in within-category variation between Polish and French), it is possible that we simply did not observe the difference with the speakers analyzed here. However, this seems unlikely given that there was a significant difference in amount of variation between Polish /ʂ/ and French /ʃ/ in

[9]This would crucially depend on obtaining data from Polish speakers who have not merged the sibilants and still make the three-way place distinction in this context. If the speakers only contrast two sibilants, Contrast-Dependent Variation predict no difference in within-category variation between Polish and French. See §3.1 for additional discussion of sibilant mergers in Polish.

the opposite direction which would be expected based on Lindblom (1986). The null result between the dental sibilants also cannot necessarily be used to argue that the amount of F2 variation in these languages should always same across all potential F2 dimensions, as the present analysis focuses only on F2 at 10ms into the vowel.

3.5.2 Retroflex variation in Polish

We did observe a significant difference in amount of variation between Polish /ʂ/ and French /ʃ/ where the Polish speakers exhibited more within-category within-speaker variation relative to the French speakers. There are multiple factors that might affect retroflex variation which are not related to phonological contrast. First, there is greater articulatory variability in retroflex articulation relative to the other sibilants (Bukmaier et al., 2014). This could lead to more acoustic variability in the retroflex sibilants generally, indicating that the effect of more variation in Polish is not a language-specific effect but a retroflex-specific effect.[10] However, given that the task did not directly manipulate speech rate, it is unlikely that substantial variation due to speech rate would be observed in the data here.

In addition, it is possible that sound change in progress is currently affecting the realization of the retroflex sibilant for these speakers. As described in §3.1, there are multiple dialects of Polish which merge the retroflex with either the dental or alveopalatal. The retroflex is distinct from the dental sibilant on the COG dimension for all speakers in the data here. For some speakers, there is a large degree of overlap in F2 values between the retroflex and alveopalatal, potentially indicating a partial merger or merger in progress.[11] A partial merger could result in higher

[10]The findings from Mandarin in Chapter 4 corroborate this and the similarities are discussed further there.

[11]Tokens for each of the speakers were easily perceptually distinguished by two native speaker consultants, which suggests the contrast is not completely merged for these speakers.

amounts of within-category variation relative to a speaker with three distinct and stable categories.

3.5.3 F2 and Polish sibilants

The results here showed not only differences in F2 onset/transition following the alveopalatal sibilant, but differences in F2 that persist throughout the entire vowel. This is not totally expected based on the previous literature which focuses on the importance of formant transitions. For all speakers examined here, F2 values of [ɛ] are consistently raised following the alveopalatal sibilant relative to other instances of /ɛ/. This offers additional support for the conclusion of Bukmaier and Harrington (2016), who show that the alveopalatal sibilant has a palatalizing influence on the following vowel. However, the data here suggest that this coarticulatory influence is not implemented identically across all speakers and vowels. More variation between-speakers was observed in [ɔ] and [a] with only some speakers producing raised F2 which persisted through the entire vowel and others producing differences only in transition.

There was the most consistency across speakers in raising F2 following alveopalatal sibilants in the mid front vowel [ɛ]. It is possible that this could be due to more coarticulation with the alveopalatal sibilant and the mid front vowel based on the proximity of tongue gestures. It could be the case that there is a higher degree of gestural overlap between the fricatives and the mid front vowel, resulting in more consistency across speakers in the F2 patterns in /ɛ/ relative to the other vowels.

The speakers in this study were all bilingual with English and living in the United States at the time of recording. It seems unlikely that the English experience would influence their Polish pronunciation such that they produce enhanced coarticulation between the alveopalatal sibilant and following vowels. However, it remains a possibility that the English environment has affected the acoustic correlates of the contrast

and different results would be obtained from Polish speakers living in a predominantly Polish language environment.

Nowak (2006) showed that removing the vocalic transitions made the alveopalatal and retroflex sibilants confusable in perception. The results here suggest that the entire vowel, not just the transitions into the sibilant, could play a role in perception. Specifically, the consistent differences in formant values suggest that speakers may be able to identify the preceding sibilant solely from later portions of the following vowel. The differences across vowel also suggest that speakers may utilize different cues (transition vs. steady-state F2) for perception of the contrast across different vowels. A perception study would need to be done to further clarify how the vocalic cues are used in sibilant perception.

As discussed in Section 3.1, diachronic mergers have been predicted for the Polish sibilants (and observed in some nonstandard varieties). The F2 differences in [ɛ] following the two sibilants might be analyzed as the transfer of phonological contrast to the vowel as part of a sibilant merger. Since the pattern is only consistent in [ɛ], it seems unlikely that these differences are indicative of transfer of contrast or merger in progress. However, there was one speaker who produced near-identical formant transitions for [ɔ] and [a]. This speaker appears to have a partial merger of the vocalic correlates for those vowels, but still distinguishes the sibilants with F2 values in [ɛ].

It remains a possibility that the Polish speakers could make additional distinctions between the sibilants on other phonetic dimensions which are not investigated here. For example, the speakers which seem to show a partial merger could contrast the sibilants using formants other than F2 or other information in the fricative spectra. The dimensions examined here (F2 and COG) are likely the most relevant based on the existing literature on production and perception of Polish sibilants. However,

there is additional information in the acoustic signal and investigating the involvement of other dimensions in this contrast is an area for future work.

3.6 Conclusion

The results in this chapter demonstrate a crucial difference in Contrast-Dependent Variation relative to Lindblom (1986). When using Contrast-Dependent Variation to make predictions about relative differences in amount of variation, each phonetic dimension is considered to be a "system" for evaluating the prediction. This accounts for the fact that while French does not use F2 as a primary cue to the sibilant contrasts, it does use F2 as a primary cue to the following vowel contrasts. Contrast-Dependent Variation therefore does not predict any difference between the two languages and no difference was observed here.

The Polish data also raise questions about the nature of the sibilant contrast for these speakers. Specifically, the results diverged from some previous work on Polish sibilants in that the F2 differences following the alveopalatal sibilant frequently persisted throughout the entire duration of the vowel. A similar pattern was observed with the alveopalatal sibilant in Chapter 4 in Mandarin and I discuss both cases further in §4.7.

CHAPTER 4

WITHIN-LANGUAGE CASE STUDY: SIBILANT FRICATIVES IN MANDARIN

4.1 Introduction

I have argued for a revision of Lindblom's hypothesis which is operationalized over phonetic dimensions instead of phonological inventories. This hypothesis was tested by examining relative differences in extent of variation between-languages in Chapters 2-3. In this chapter, I formulate and test an extension of the Contrast-Dependent Variation hypothesis that predicts differences in extent of variation between speakers of the same language. Given a contrast where speakers differ in which phonetic dimension serves as primary cue: We expect variation to emerge on dimension B for speakers that primarily use dimension A for contrast (further specified in §4.3).

I test this hypothesis by examining sibilants in Mandarin, where speakers show individual differences in how the sibilant contrasts are realized in phonetic space. Using similar experimental methodology to Chapter 2, I compare the degree of contrast in spectral center of gravity (COG) with the amount of variation in the onset of the second formant of the following vowel (F2). The main finding is that amount of F2 variation increases with degree of COG contrast across speakers. Implications for perception and cue weighting are also discussed.

Table 4.1. Consonant inventory of Mandarin (Duanmu, 2007)

	Labial	Alveolar	Retroflex	Alveopalatal	Velar
Stop	p p^h	t t^h			k k^h
Affricate		ts ts^h	ʈʂ $ʈʂ^h$	tɕ $tɕ^h$	
Fricative	f	**s**	**ʂ**	**ɕ**	x
Nasal	m	n		ɲ	(ŋ)
Approximant		l			

4.2 Background

4.2.1 Mandarin sibilants

Mandarin has a rich fricative inventory with five places of articulation for fricatives and affricates. Mandarin sibilants have been characterized as having a variety of different places of articulation: dental, denti-alveolar, retroflex, laminal post-alveolar, and apical post-alveolar. Chang and Shih (2015) provides a review of these claims; some of this variation is likely attributable to data collection in different regions. In this study, I use the terms dental, retroflex and alveopalatal (Ladefoged and Wu, 1984; Duanmu, 2007; Chang and Shih, 2015). The existence of three sibilant categories and variation in phonetic implementation across speakers is of importance for testing the hypotheses here and the exact places of articulation are not crucial. The consonant inventory of Mandarin is given in Table 4.1.

Acoustically, the three sibilants have sometimes been described as having a three-way contrast in spectral center of gravity[1] (COG; Lee, 1999; Lee-Kim, 2011; Kallay and Holliday, 2012). Other studies have reported a two-way center of gravity contrast between the alveolar and the other two sibilants and an F2 onset contrast distinguishing the alveopalatal from the retroflex (Stevens et al., 2004). COG has also been shown to be influenced by coarticulation with following vowels—COG of alveolar and retroflex sibilants is lower when followed by a rounded vowel and a smaller

[1]Spectral mean is a term also used in the literature and is synonymous with spectral center of gravity. I use COG here.

COG difference between the alveolar and retroflex has been found before /u/ relative to other vowels (Jeng, 2006; Li, 2009). However, Hu (2008) examined articulation and acoustics and found that speakers, not vowel context, were the major source of variability in both articulation and acoustics. As I am focusing on the effect of COG, I only examine the sibilants in Mandarin. Mandarin does have other fricatives for which COG is likely a relevant cue in perception and production, however COG is not expected to be a primary cue distinguishing non-sibilant fricatives (cf. Jongman et al., 2000).

There is variation across regional dialects in realization of the sibilants. It is sometimes claimed that Taiwan Mandarin and other southern dialects lack the contrast between the alveolar and the retroflex sibilant (Kubler, 1985; Duanmu, 2000). However, some work shows only a partial merger–while the contrast might be less distinct in Taiwan Mandarin, various factors influence degree of contrast in the dialect. Vowel context, sociolinguistic factors, formality of task, contrastive focus, and other types of prosodic prominence have all been shown to enhance the alveolar-retroflex contrast even for Taiwan speakers (Chung, 2006; Jeng, 2006; Li, 2009; Chuang and Fon, 2010; Chang and Shih, 2012, 2015).

It is difficult to determine exactly how widespread the alveolar-retroflex merger (or partial merger) is given that most of the work on the merger has focused only on Taiwan Mandarin. The use of retroflexion also has socio-indexical value; it is associated with higher education levels and distinguishes standard Mandarin pronunciation from "dialect-accented" Mandarin (Chang et al., 2013). Given the sociolinguistic situation, it is possible that many speakers who may have the merger in casual contexts distinguish the sounds fully in formal and/or laboratory contexts.

In perception, several studies have found the primary cue for the retroflex-alveolar contrast to be COG or the position of the lowest spectral prominence (Wu and Lin, 1989; Li, 2008; Chang, 2013). Li (2008) argues that COG is not sufficient to distin-

guish the alveopalatal from the other two sibilants and the primary cue distinguishing /ɕ/ is instead F2 onset of the following vowel. There is also dialectal variation in perception of the fricatives. Chang (2013) compared perception between Taiwan Mandarin (alveolar-retroflex merger) and Beijing Mandarin (no merger) listeners and found different perceptual boundaries for cross-dialectal perception.

There is an allophonic restriction on sibilants requiring [ɕ] before high front vowels (Duanmu, 2007; Lin, 2014). Because of this positional neutralization, some have argued that the alveopalatals can be represented as underlying velars which become palatalized before high vowels (Wu, 1994). This matches the diachronic evidence which suggests that the alveopalatals arose in Mandarin due to velar palatalization (Chao, 1965; Li, 1999). However, all sibilants contrast preceding the vowels [a] and [əu u] synchronically (Duanmu, 2007; Li, 2008; Lin, 2014). Therefore, all three sibilants are considered to be independent phonemes in many synchronic analyses (Li, 1999; Cheng, 2011) and are assumed to be distinct phonological categories in this study.

In sum, the previous work on Mandarin sibilants has shown variation in how the sibilant contrasts are realized phonetically among individual speakers and regional dialects. Specifically, some acoustic studies report data from speakers with a three way COG contrast between /s ʂ ɕ/, while others report relatively more use of F2. There is a also merger between the alveolar and retroflex which is common in some regional dialects and often associated with lower social prestige.

4.2.2 Cue weighting in production

In this chapter, I investigate the relationship between degree of contrast on the COG dimension and extent of within-category variation on the F2 dimension. I take "degree of contrast" to mean the relative strength of a particular phonetic dimension

in predicting phonological category. This is typically referred to as cue weight[2] and is often calculated using classification algorithms. In the data here, cue weights differ across speakers such that not all speakers use the same dimension as the primary cue to the sibilant contrasts.

Cue weight in production is typically measured using a classification algorithm (e.g., discriminant analysis, logistic regression) which assigns relative weights to each predictor dimension. Differences in cue weighting patterns for the same contrast have been observed in production between native speakers of the same language (Shultz et al., 2012), native and non-native speakers (Schertz et al., 2015), non-native speakers with different levels of L2 exposure (Kong and Yoon, 2013), and speakers of a language undergoing sound change (Bang et al., 2018; Coetzee et al., 2018; Kuang and Cui, 2018).

In this chapter, I investigate the question of whether cue weight of COG is predictive of variation in F2. Previous work on individual variation in cue weights has found a related (but non-identical) correlation. Shultz et al. (2012) examined native English stop production and found an inverse correlation between weight of VOT and weight of F0 across speakers, indicating that speakers who used the VOT dimension more contrastively used the F0 dimension less contrastively. The experiment in this chapter builds on this work by examining the relationship between cue weight and variability. Implications for cue weighting are further discussed in §4.7.

4.3 Predictions

In Table 4.2, I compare two instances of Contrast-Dependent Variation that relative differences in within-category phonetic variation can be predicted by differences in phonological contrast implementation. The between-language prediction was tested

[2]Here I am referring only to cue weight in production. See Chapter 1 for additional discussion of cue weighting in production vs. perception.

Table 4.2. Between vs. within-language predictions of Contrast-Dependent Variation

Prediction	**Summary**
Within-languages	Given a phonological contrast with two phonetic dimensions X and Y serving as cues and between-speaker variation in which dimension is used as the primary cue: we expect relatively more within-category within-speaker variability in X for speakers who show relatively more contrast on Y.
Between-languages	For a given phonetic dimension X: we expect less group-level within-speaker variability and less between-speaker variability in languages which employ X as a primary cue to a phonological contrast relative to languages which do not employ X as a primary cue to a phonological contrast.

in Chapters 2-3, the within-language prediction is tested here. See Chapter 1 for detailed definitions of relevant terms in these hypotheses.

In the case of the Mandarin sibilants, we consider a 3-way phonological contrast between /s ʂ ɕ/. While there are many phonetic dimensions potentially involved in this contrast, previous literature points to COG and F2 as being the two dimensions which are used as primary cues in production and perception. There is between-speaker variation in how many sibilants are contrastive and which dimensions are used as primary cues for each contrast (as discussed in §4.2). Given this situation, we expect more variation on the dimension which is not the primary dimension for contrast. Specifically, we expect relatively more variation in F2 in speakers who use COG more to distinguish the sibilant contrasts (operationalized as cue weight in production).

In Figure 4.1, I provide schematics showing the expected between-speaker differences predicted by Contrast-Dependent Variation. Speaker A (top panel) uses COG as the primary cue for all three sibilant contrasts and exhibits more variation in the F2 dimension relative to Speaker B (middle panel) who uses COG as the primary cue for only two of the sibilant contrasts. If this pattern holds across speakers, we

expect these between-speaker differences to amount to positive correlations between F2 variation and COG contrast when compared across speakers. Schematic plots showing the expected outcome are given in the bottom panel of Figure 4.1. If COG contrast and F2 variation are related in the way predicted here, we should expect to see a positive correlation between these values across speakers.

4.4 Experimental design

The methods were matched as closely as possible to the methods from Chapter 2 (described in detail in §2.3), with some necessary changes. All parts of the experiment were conducted in Mandarin by native speaker research assistants.

4.4.1 Participants

All speakers were between the ages of 18-30 and recruited at The University of Massachusetts Amherst and surrounding colleges. Most speakers were undergraduate students. Participants were recruited through the Linguistics Department's participant recruitment system and through email advertisements to the Taiwanese and Chinese Students' Association. All recruitment materials (emails, sign-up info, etc.) were distributed in Mandarin orthography.

19 Mandarin speakers were recorded. All speakers acquired Mandarin natively in China and relocated to the United States for college or high school. One speaker was excluded because they did not complete the task. One speaker was excluded due to frequent speech errors (difficulty on more than 25% of stimuli). Additional noise from a room fan and/or use of breathy voice prevented accurate formant tracking for six speakers. After these exclusions, data from 11 speakers were fully analyzed.

4.4.2 Stimuli

The stimuli in Mandarin were words and rare words which we expect to behave as non-words. Because the Mandarin writing system is logosyllabic, developing new

Figure 4.1. Expected results under Contrast-Dependent Variation. Top panel: Predicted speaker with relatively more COG contrast and more F2 variation. Middle panel: Predicted speaker with less COG contrast and less F2 variation. Bottom panel: Predicted relationship between COG contrast and F2 variation across speakers.

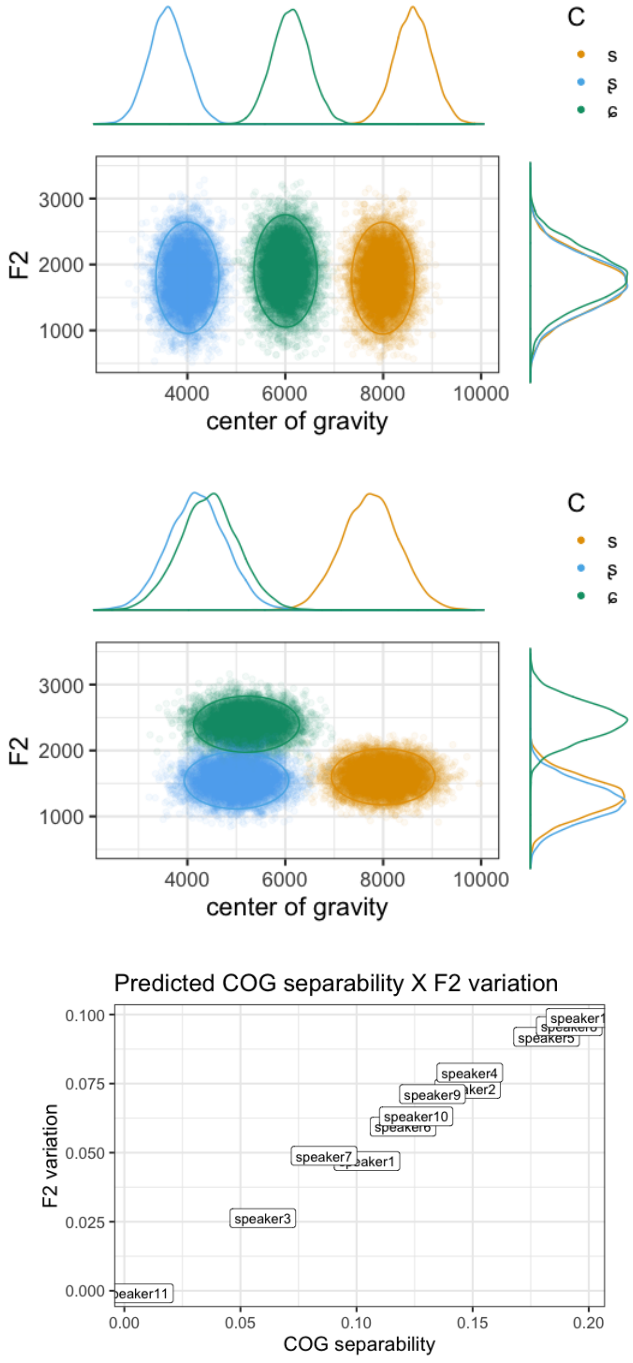


Figure 4.2. An example prompt screen from the Mandarin experiment.

symbols for non-words presents several problems for participant reading. Instead of attempting to design new and orthographically natural characters, we used rare words with existing characters as "non-words". Each stimulus was presented with the simplified Mandarin orthographic character and the pinyin script, a romanized quasi-phonemic orthographic system. With the pinyin presented alongside the logosyllabic characters, the participants were able to pronounce the intended stimulus even if they were unfamiliar with the word or Mandarin character. No participants self-reported trouble reading either orthographic system. The stimuli were read in the carrier phrase "wǒ bǎ X dú yī biàn" ('I read X once'). An example prompt screen is shown in Figure 4.2.

Mandarin stimuli were crossed according to the following factors: sibilant (3 levels: s ʂ ɕ) × vowel context (3 levels: i a u) × word status (3 levels: high frequency/low frequency/non-word) × number of syllables (2 levels) × tone (4 levels). Word frequency judgments were provided by two native speaker research assistants for the initial frequency classifications. The participants were also given a frequency judgment task to verify the research assistants' intuitions. Not all factors could be fully crossed: there is a phonotactic restriction that requires the alveopalatal sibilant be-

Table 4.3. Example Mandarin stimuli

Language	C	vowel	stimulus (IPA)	stimulus (pinyin)	tone	frequency
Mandarin	s	a	sá	sa	4	low
Mandarin	s	a	sā	sa	1	high
Mandarin	s	a	sá	sa	4	rare
Mandarin	ʂ	a	ʂā	sha	1	low
Mandarin	ʂ	a	ʂǎ	sha	3	high
Mandarin	ɕ	a	ɕā	xia	1	low
Mandarin	ɕ	a	ɕà	xia	2	high
Mandarin	ɕ	a	ɕá	xia	4	rare
Mandarin	s	u	sú	su	4	rare
Mandarin	s	u	sū	su	1	high
Mandarin	s	u	sǔ	su	3	low
Mandarin	ʂ	u	ʂū	shu	1	low
Mandarin	ʂ	u	ʂǔ	shu	3	high
Mandarin	ʂ	u	ʂù	shu	2	rare
Mandarin	ɕ	u	ɕǔ	xiu	3	low
Mandarin	ɕ	u	ɕu	xiū	1	high
Mandarin	ɕ	a	ɕá	xia	4	rare
Mandarin	ɕ	i	ɕī	xi	1	low
Mandarin	ɕ	i	ɕì	xi	2	high
Mandarin	ɕ	i	ɕí	xi	4	rare

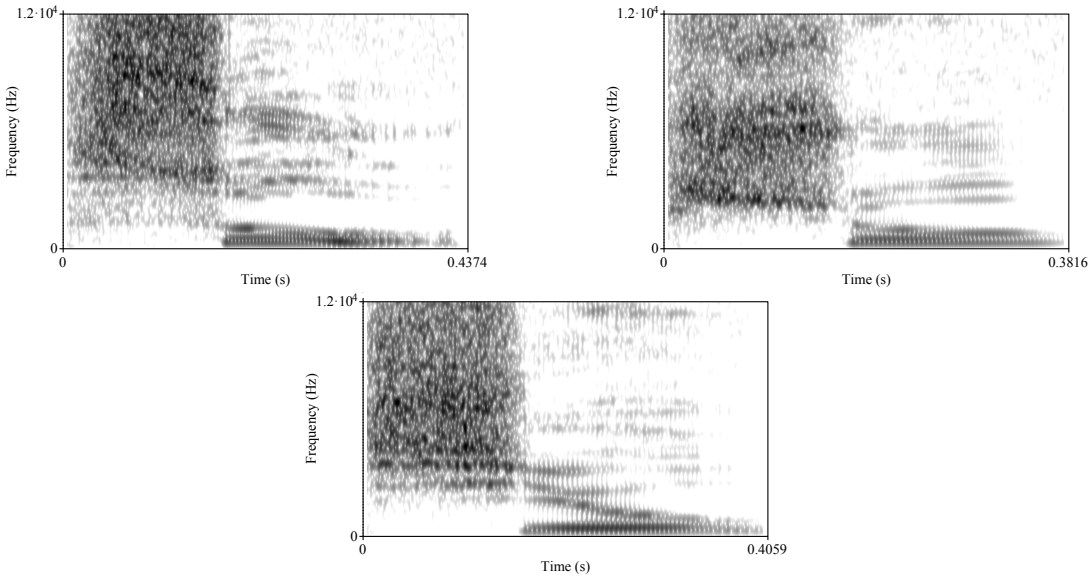
fore [i], so the three sibilants are only fully crossed in the [a] and [u] contexts. Due to limitations of the lexicon, some of the tones are not fully crossed with all other factors. There were a total of 137 distinct sibilant stimuli.[3] Additional stimuli with word-initial affricates and stops were included as fillers. Word-initial non-sibilant fricatives were not included in the task.

4.4.3 Recording

Recording was done according to the same procedure used in Chapter 2. The only difference was the inclusion of additional stimuli in the task. Stop and affricate tokens were elicited along with the voiceless sibilants, neither of which are analyzed

[3]The full stimuli list is available in the data archive for this project at https://osf.io/2famr/.

Figure 4.3. Alveolar, retroflex, and alveopalatal sibilants in Mandarin. Left: /su/. Right: /ʂu/. Bottom: /ɕu/.



here. After recording the participants did the word frequency judgment task (as in Chapter 3) to ensure the rare words were actually unknown to the participants.

4.4.4 Data processing and analysis

The data processing followed similar methods to those used in Chapters 2-3. The recordings from each speaker were first scanned by the author and research assistants for speech errors. The recordings were forced aligned using the Montreal Forced Aligner McAuliffe et al. (2017) using a pretrained Mandarin model.[4]

A Praat script based on DiCanio (2013) was used to extract spectral moments of the fricatives and formant values of the following vowels. The spectral moments were time-averaged over the middle 60% of the fricative interval. The formants were estimated using the Burg method and extracted at 10 ms intervals throughout the duration of the vowel. Formant excursions greater than 1000 Hz over 10 ms were

[4]Available at https://montreal-forced-aligner.readthedocs.io/en/latest/pretrained_models.html.

assumed to be tracking errors and were excluded. This excluded a total of 28 observations across all speakers, sibilants, and vowel contexts.

Example tokens are shown in Figure 4.3. As expected based on the previous literature, the alveolar sibilant exhibits a higher center of gravity relative to the other two sibilants and the alveopalatal exhibits higher F2 relative to the other sibilants.

4.5 Results: Differences in contrast implementation

In accordance with the previous literature, we found differences across speakers in use of COG vs. F2 in the sibilant contrast. In this section, I show example speakers with different phonetic implementations of the sibilant contrast. Graphs for all speakers are included in Appendix C.1.

4.5.1 Contrasts in COG

All speakers distinguished the alveolar sibilant from the other two sibilants in COG. There was variation between-speakers in degree of COG overlap between the retroflex and alveopalatal, with some speakers having almost total category overlap and other speakers having little category overlap. In Figure 4.4, I show two example speakers which appear to have three distinct sibilant categories on the COG dimension. Compare these with the example speakers shown in Figure 4.5, which appear to distinguish only two of the sibilants in COG alone. Speaker m-02 has almost total overlap between the retroflex and alveopalatal sibilants in COG. Speaker m-15 appears to have two distinct categories between the alveolar and retroflex, while the alveopalatal COG values overlap with both categories.

4.5.2 Contrasts in F2

There was consistency across speakers in use of higher F2 following the alveopalatal sibilant. These differences consistently persist throughout the entire duration of the following vowel in all vowel contexts. There was between-speaker variation in mean

Figure 4.4. Example speakers: 3 distinct categories on COG dimension. COG in Hz.

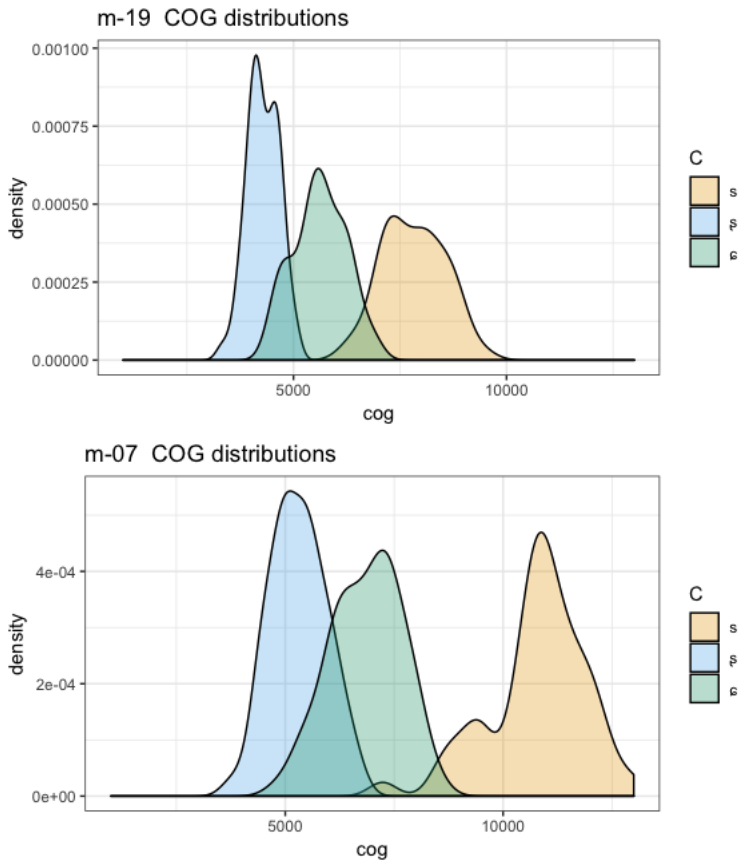
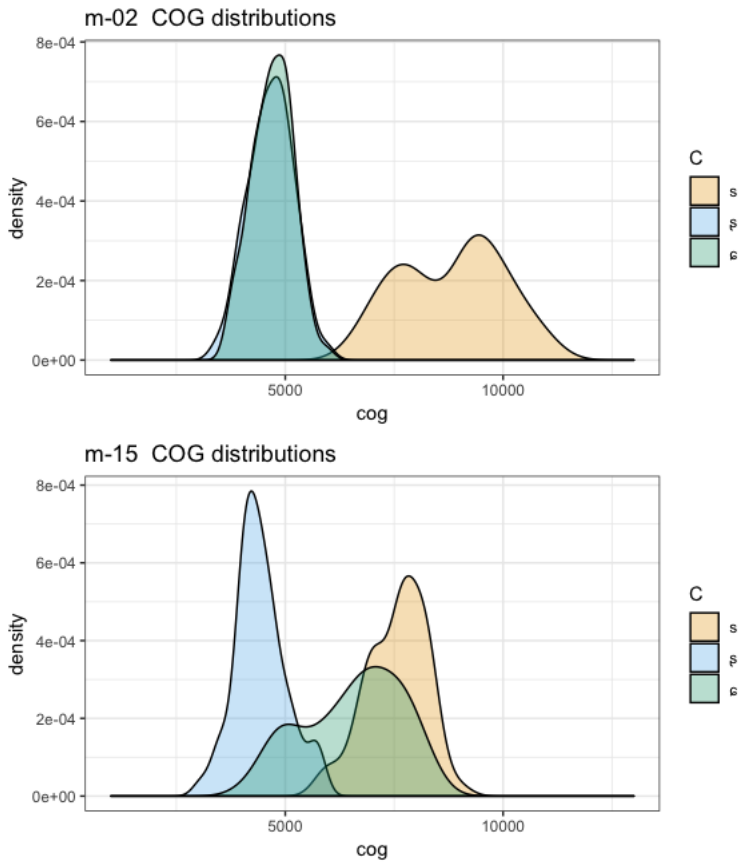


Figure 4.5. Example speakers: 2 distinct categories on COG dimension. COG in Hz.



F2 values and within-category variance. There were also differences in group-level within-speaker variation between the /a/ and /u/ contexts, with speakers exhibiting generally more within-category variation in F2 for sibilants preceding /u/.

Representative speakers are shown in Figures 4.6-4.7. These speakers reflect the consistent group pattern of higher F2 values throughout the duration of the vowel following the alveopalatal sibilant and more within-speaker variation in /u/ relative to /a/.

These results are similar to the results for Polish (Chapter 3), which exhibits a similar sibilant contrast. The vowels following the alveopalatal sibilant in Polish also exhibit raised F2 values, with differences frequently extending the full duration of the vowel (though this differed somewhat across vowel contexts). The consistency of this finding across these two languages suggests a general coarticulatory effect of the alveopalatal sibilant. I return to this issue in §4.7.

4.5.3 Contrasts in the two dimensional space

The figures in 4.8-4.11 show the three sibilant contrasts in a two-dimensional COGxF2 space for example speakers. These speakers are examples from the group and are intended to reflect the between-speaker variation in contrast implementation which is present in the data. Figures showing data from all speakers are given in Appendix C.1.

The F2 values shown here are the F2 onset measurements at 10 ms into the following vowel. Density plots are given along the x and y axes which show the distributions of the tokens in the COG and F2 spaces respectively. Each dot in the main panel represents the values from a single token in the two-dimensional space.

In Figure 4.8, I show a speaker who seems to have three distinct distributions on the COG dimension and 2 distinct distributions on the F2 dimension. This is

Figure 4.6. Example speakers: Formant trajectories of /a/ following the three sibilants. F2 in Hz.

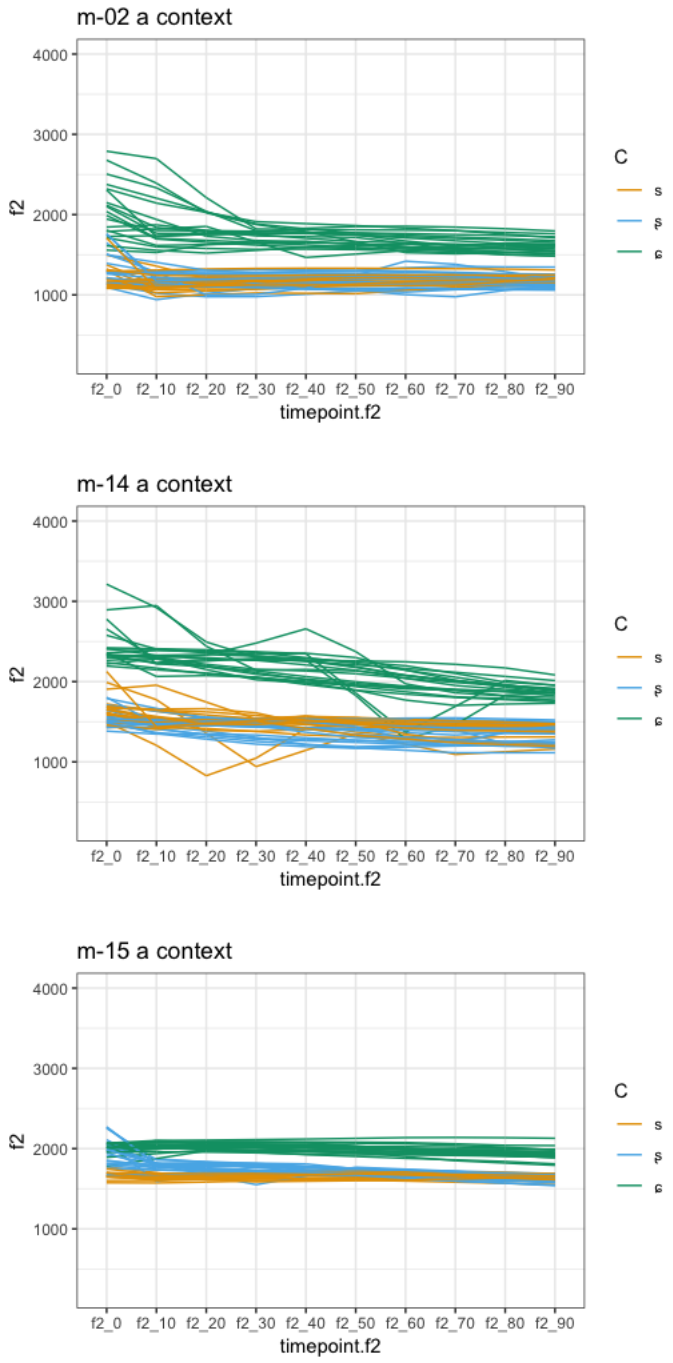


Figure 4.7. Example speakers: Formant trajectories of /a/ following three sibilants. F2 in Hz.

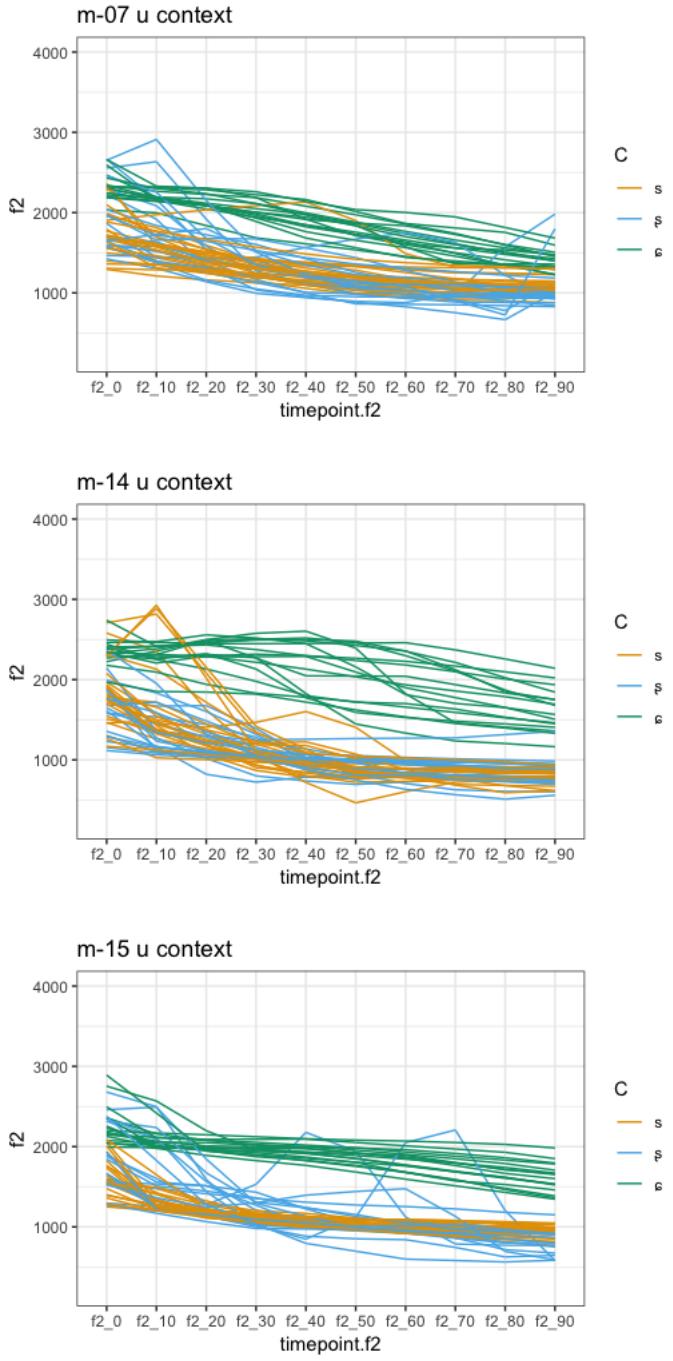
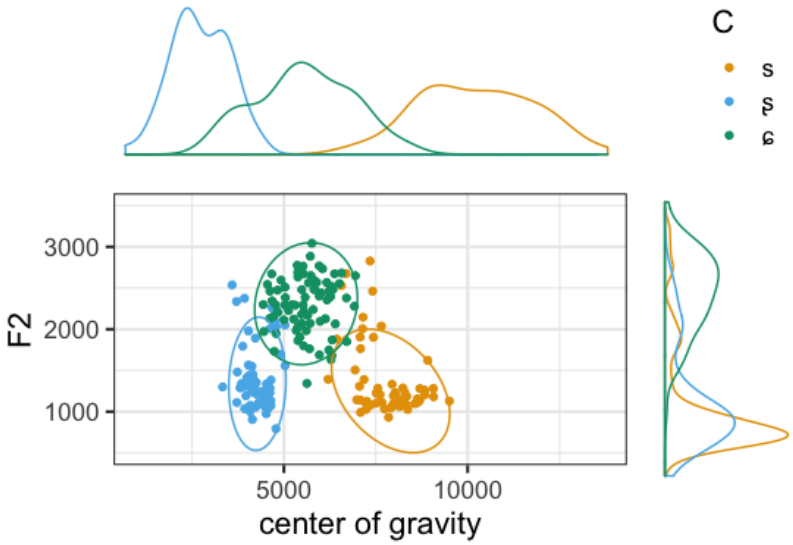


Figure 4.8. Sibilant contrasts in phonetic space: Speaker 19. F2 and COG in Hz.



the speaker with greatest COG distinction between categories, although two other speakers show a similar pattern with slightly more category overlap.

In Figure 4.9, we see a speaker who has two distinct distributions on both the COG and F2 dimensions, but three distinct categories in the two dimensional space. The speaker in Figure 4.10 also has two distinct distributions on both dimensions but the alveopalatal shares similar COG values with the alveolar instead of the retroflex as with speaker 02 in Figure 4.9. Several other speakers are similar in that they have 3 distinct categories in the two-dimensional space, but have higher degrees of overlap one each individual dimension.

Out of 11 total speakers, two speakers appeared to merge the alveolar and retroflex sibilants, despite none of the speakers being from regions typically associated with the merger. A speaker with a potential merger is shown in Figure 4.11. This speaker shows some separability between categories, but has a large degree of overlap between the retroflex and the alveolar. Previous literature (see §4.2 for a review) suggests that in case of the alveolar-retroflex merger, it is typical for phonologically retroflex tokens to be realized as alveolar. However, the speaker in 4.11 seems to display the opposite

Figure 4.9. Sibilant contrasts in phonetic space: Speaker 02. F2 and COG in Hz.

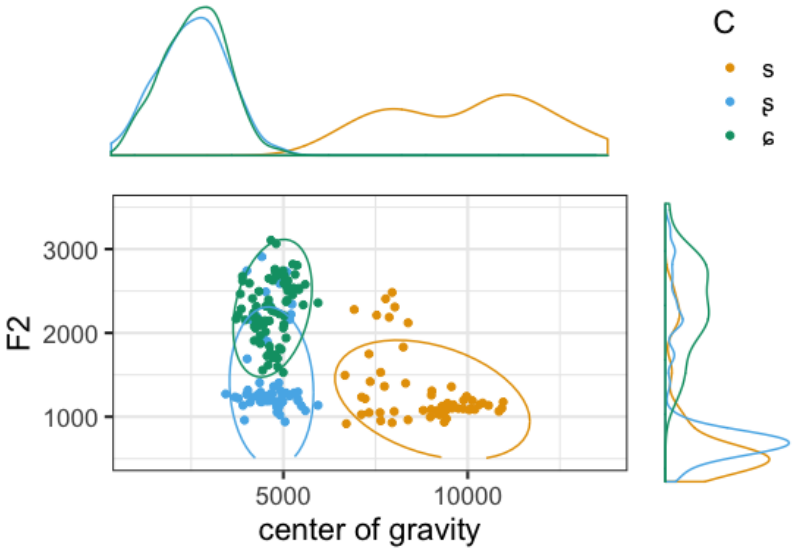


Figure 4.10. Sibilant contrasts in phonetic space: Speaker 06. F2 and COG in Hz.

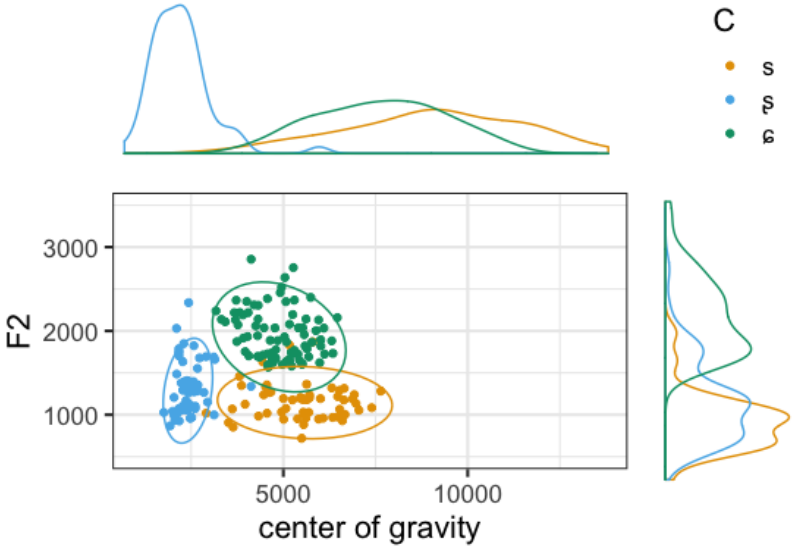
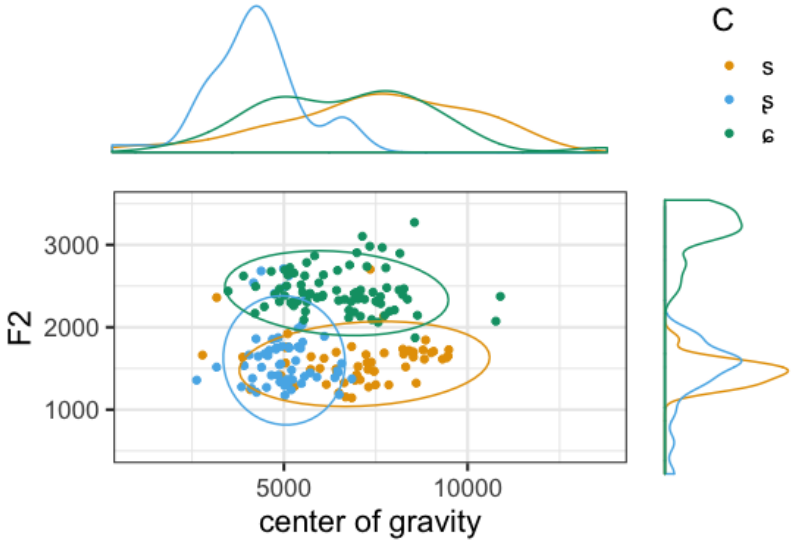


Figure 4.11. Sibilant contrasts in phonetic space: Speaker 08. F2 and COG in Hz.



pattern. This speaker exhibits more variation in the alveolar category, with lower COG values overlapping with the retroflex distribution.

4.5.4 Interim discussion: Differences in contrast implementation

These results show individual differences in how the sibilant contrasts are instantiated in the phonetic space of F2xCOG. There are no speakers which seem to use the COG or F2 dimensions exclusively and all speakers consistently exhibit higher F2 values for the alveopalatal sibilant. Even speakers which show 3 distinct categories along the COG dimension (as in Figure 4.8) still show relatively higher F2 values following /ɕ/.

Most speakers appear to have the 3-way contrast in this space, (only two speakers appear to have a potential merger between /s/ and /ʂ/). It could be the case that more of the speakers would show a merger in natural speech. Given the socio-indexical value of /ʂ/ in Mandarin (see §4.2), it is possible that speakers who would have the merger in natural speech produced retroflexion in the lab setting due to prescriptive influence or as part of a hyperarticulation strategy. Further work would need to be done to clarify the effect of the laboratory setting on retroflex production. I discuss

the retroflex merger further in §4.7. For the purposes of this analysis, whether any individual speaker has the merger (and is what context) is not of critical importance. All speakers, including the ones who appear to merge /s ʂ/, can be analyzed and compared using the same methodology.

4.6 Results: Effect of contrast on variability

4.6.1 Quantifying variables

In order to test the hypothesis that we expect relatively more F2 variation in speakers who use the COG dimension more for contrast, we need to quantify both F2 variability and COG contrast. F2 variance was calculated within speaker, sibilant, and vowel context. These variances were divided by the within category mean to calculate the coefficient of variation (which is unitless). This was done to abstract over differences in mean values and compare the variation across speakers and contexts.

4.6.1.1 Quantifying cue weight with LDA

COG contrast was quantified with Linear Discriminant Analysis (LDA; Fisher, 1936; Fukunaga, 1990; Duda et al., 2012). LDA is a classification method that relates continuous predictor variables to category labels. Classification is achieved through finding the linear combination of predictor features which best separates two categories.[5] LDA assumes the category distributions are drawn from Gaussian distributions with a common covariance matrix.

The purpose of LDA is to find the linear function that can best discriminate a set of categories (here the sibilant categories) given a set of predicting features (here the acoustic measures of COG and F2). LDA is often used as a dimensionality reduction technique to find linear combinations of existing features which maximally

[5]LDA therefore only determines *linear* separability. Similar discrimination algorithms can be applied which utilize different functions for separability (e.g., quadratic discriminant analysis). The methodology for using these would be similar though they are not directly examined here.

separate the relevant categories. This is common in the literature on automatic speech recognition (see Haeb-Umbach and Ney, 1992; Viszlay et al., 2012, for a review).

There is a precedent in the phonetics literature of using LDA for the purpose of determining cue weight in production (Shultz et al., 2012; Garellek and White, 2015; Schertz et al., 2015; Kim and Clayards, 2019, among others). There are many classification and feature selection algorithms that could potentially be used to classify phonetic observations into categories. However, I focus on LDA for the analysis here as methodological consistency allows us to compare the present results to previous results on cue weighting in production (see §4.7 for such comparisons). I provide further discussion of alternative analyses in Appendix C.2.

Following the previous literature in phonetics, I use the coefficients of linear discriminants as the measure of cue weight from LDA. The coefficients are regression weights used to calculate the probability of category membership (James et al., 2013). They indicate the contribution of each predictor variable to the discriminant function; higher values indicate more contribution to the discriminant function.[6] These weights can be interpreted as indexing the strength of each individual predictor.

An LDA was performed in each vowel context for each speaker using COG as the relevant predictor. In the results that follow, I take the coefficients of linear discriminants to be the cue weights, the quantitative measure of contrast on the COG dimension (see §4.6.1.1 for background on LDA methodology).

4.6.2 Correlations across speakers

In Figure 4.12, I show the COG coefficients and F2 variability values partitioned by sibilant and vowel context. We can see differences in overall F2 variation according to

[6]The polarity of the coefficients will depend on the coding of factors. High positive values and low negative values both indicate high contribution to the discriminant function. Coefficients in phonetics are therefore sometimes presented with reverse polarity such that higher values always indicate more weight regardless of how factors are coded (Shultz et al., 2012; Schertz et al., 2015).This is also done here when appropriate.

sibilant and vowel context by comparing the values along the y axes of the individual panels. There does seem to be a positive relationship across speakers for [s] in both vowel contexts. However, there is generally less variation in the [a] context relative to the [u] context, which can also be seen in the raw formant trajectories (Figures 4.6-4.7. There is also one speaker, m-15, who exhibits low variability in F2, yet has high COG coefficients.

The results for the alveopalatal sibilant also appear to show positive correlations in both vowel contexts. Unlike in the alveolar and retroflex results, there does not seem to be a general difference in amount of F2 variation between the two vowel contexts. The results for the retroflex sibilant do appear to show a difference in amount of F2 variation between the vowel contexts but do not appear to show a positive correlation between F2 variation and COG contrast.

4.6.3 Modeling F2 variation effects with regression

The hypothesis is that speakers who exhibit relatively more COG contrast should also have relatively more F2 variation. In order to quantify the effect of COG contrast, I use a linear mixed effects regression model to predict differences in F2 variation. The dependent variable is F2 variation within-speaker, sibilant category, and vowel context. The factors included in the model are: COG contrast (COGcoefs; as measured by the LDA coefficients), sibilant (C), vowel context (V), COG coefficients $\times$ sibilant interaction, and random intercepts for speaker.

This means that we expect a significant effect of the COG coefficients in the regression output, which shows the effect of COG contrast in the intercept case (here /ʂa/). If the same relationship between COG contrast and F2 variation holds across all the sibilants we expect non-significant interactions between the COG coefficients and each sibilant. Significant interactions between the COG coefficients and an indi-

Figure 4.12. COG coefficients and F2 variation across speakers: Alveolar sibilant in top panel, alveopalatal sibilant in middle panel, retroflex sibilant in bottom panel. F2 variation is the unitless coefficient of variation.

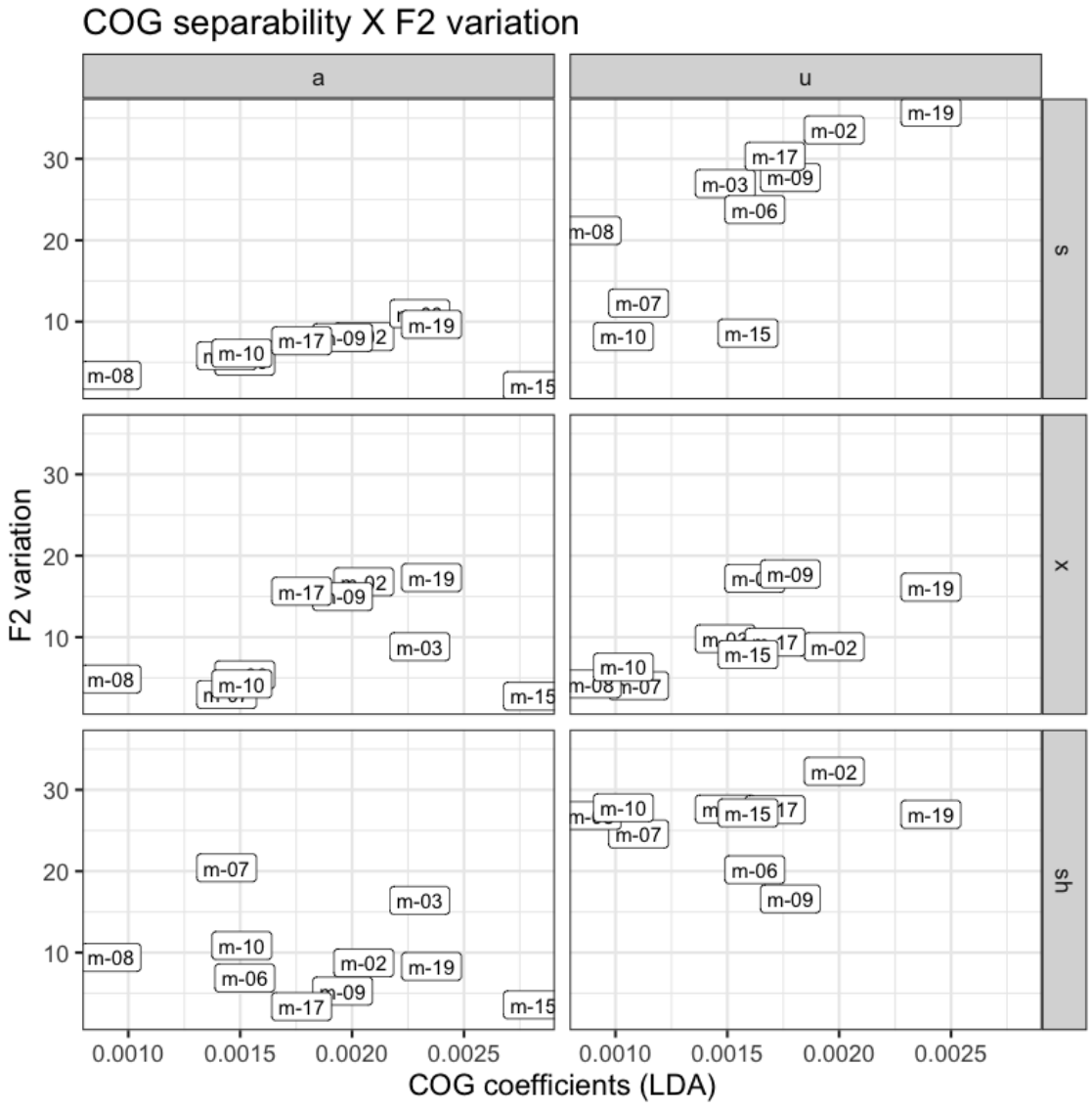


Table 4.4. Fixed effects table for linear mixed effects regression. Dependent variable: within-category within-vowel F2 variation, Predictors: COG coefficients, C, V, C×COGcoefs interaction, random intercepts for speaker. Intercept is [ʂa].

Fixed effects	Estimate (se)	t	p
(Intercept)	15.98(5.65)	2.83	0.007 **
COGcoefs	−2.66(3.03)	-0.88	0.384
C-/s/	−16.87(7.15)	-2.36	0.022*
C-/ɕ/	−26.32(7.15)	-3.68	< 0.001***
V-/u/	12.23(1.75)	6.99	< 0.001***
COGcoefs:C-/s/	8.21(4.01)	2.05	0.046*
COGcoefs:C-/ɕ/	10.83(4.01)	2.70	0.009**

vidual sibilant indicates that the relationship between COG contrast and F2 variation is different for that sibilant relative to the sibilant which is the intercept.

The output of the regression model is given in Table 4.4. The main effect of the COG coefficients is non-significant, indicating no significant relationship between COG contrast and F2 variability for intercept /ʂa/. The significant effects of the sibilants /s/ and /ɕ/ indicate that there is a significant difference in amount of F2 variation relative to /ʂ/. The negative values of the estimates indicate that /s/ and /ɕ/ exhibit significantly less within-category F2 variation relative to /ʂ/. The significant effect of the vowel /u/ indicates that there is significantly more F2 variation in the /u/ context relative to the /a/ context.

The crucial results are the interactions between the COG coefficients and the individual sibilants. The significant effect of COGcoefs × /s/ indicates that the relationship between COG contrast and F2 variation is significantly different for /s/ relative to /ʂ/. The positive estimate indicates that there is a positive relationship between COG contrast and F2 variation for /s/, unlike for /ʂ/ where there is no significant relationship. There is also a significant interaction for the COG coefficients × /ɕ/, indicating a positive relationship between COG contrast and F2 variation for the alveopalatal as well. In sum, these results indicate that there is overall less within-

category F2 variation in /s/ and /ɕ/ relative to /ʂ/, and this variation increases with COG cue weight for /s/ and /ɕ/ but not /ʂ/.

4.7 Discussion

The results for the alveolar and alveopalatal sibilants were in accordance with the within-language prediction of Contrast-Dependent Variation: speakers who exhibited greater contrast on the COG dimension (as measured by the LDA coefficients) also exhibited greater amounts of within-category variability on the F2 dimension. This hypothesis is an extension of Lindblom's (1986) hypothesis, which did not make any predictions about differences in variation among speakers of the same language. As I have implemented the between-language prediction to refer to phonetic dimensions instead of phonological inventories, the hypothesis can extend to predict variation differences among speakers with different phonetic realizations of the same phonological contrast. In the case study presented here, variation on the F2 dimension increased as contrast on the COG dimension increased. The effect was observed across speakers except in the retroflex sibilant.

4.7.1 Lack of effect for the retroflex sibilant

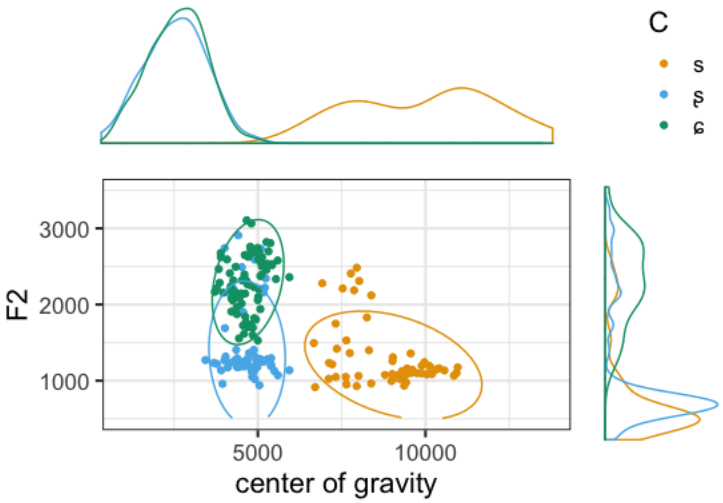
F2 variation for the retroflex category did not correlate with COG contrast across speakers. There are several reasons why this might be the case, as there are multiple factors potentially influencing retroflex variation which are independent of COG contrast and unique to the retroflex. The retroflex sibilant also behaved uniquely in the Polish results in Chapter 4, where we observed more within-category variation in retroflex production relative to the other sibilants in Polish and French. In the Mandarin data, we did not observe more within-category variation in retroflex realization generally, but we did observe more retroflex variation in the /u/ vowel context relative to the other sibilants.

First, the retroflex sibilant is involved in a merger in many dialects of Mandarin. Most of the speakers in this study produced three distinct sibilant categories in the F2xCOG space. This was expected as none of the speakers are from regions typically associated with the merger. However, there were a few speakers that displayed a good amount of overlap between the retroflex and the alveolar categories in F2xCOG space. It is also possible that there are additional speakers who merge those categories in natural speech (producing all retroflex/alveolar tokens as alveolar) but produced retroflexion in the lab due to the social prestige associated with retroflexion. If speakers realize /ʂ/ tokens as alveolar, even only in some contexts, this could result in additional within-category variability for /ʂ/. Increased retroflex variability could potentially obscure any relationship between COG contrast and F2 variation.

As in Polish, articulatory variability in retroflex production has been reported in Mandarin. In an MRI study of Mandarin sibilant production, Proctor et al. (2012) found more articulatory variation in retroflex production relative to the other sibilants. It is possible that differences in articulation independently contributed to within-category retroflex variation, leading to larger amounts of variation in general thus obscuring any contrast effects.

There is an additional factor which might constrain retroflex variation as opposed to contributing additional variation. For many speakers the distribution of the retroflex sibilant is unique in that it is bounded in the COGxF2 space. The retroflex frequently contrasts with the alveolar in COG and the alveopalatal in F2. An example speaker showing this configuration is shown in Figure 4.13. For the speaker in this figure, producing more variation along the F2 dimension in the retroflex category would result in increased category overlap with the alveopalatal. Similarly, producing more variation along the COG dimension would result in increased category overlap with the alveolar. It is possible that this constrains retroflex variation such that variation does not increase as COG contrast increases.

Figure 4.13. Example speaker with retroflex category bounded in phonetic space. F2 and COG in Hz.



The location of the mean category values as in Figure 4.13 is similar to the distribution of the Polish sibilants in Chapter 3. In that data, the retroflex sibilant often shared similar mean COG values with the alveopalatal and shared similar F2 values with the alveolar (the same configuration has been reported in previous work on Polish as well, see Chapter 3 for a review).

In sum, there are multiple factors that might independently contribute to within-category retroflex variation in addition to any phonological contrast effects. The acoustic position of the retroflex sibilant relative to the other sibilant categories, the potential involvement in a merger, and articulatory variability are all factors which could be contributing to retroflex acoustic variation. Realization of the retroflex sibilant generally exhibited more group-level within-speaker variability and did not show any contrast effects. The findings for the retroflex sibilant in Polish were similar, suggesting that these effects may not be language-specific and are likely more general to retroflex sibilant production.

4.7.2 F2 in Mandarin sibilants

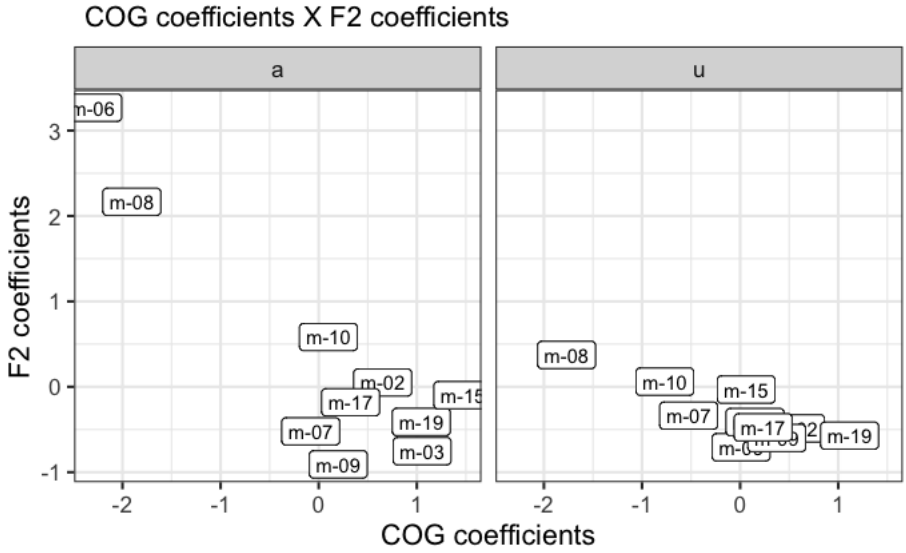
In the data here, all speakers exhibited raised F2 values following the alveopalatal sibilant which continued throughout the duration of the vowel. This is similar to what was observed for vowels following Polish sibilants in Chapter 3, though the Mandarin data show more consistency across speakers. This raises questions for perception. Specifically, can speakers identify the preceding sibilant solely from the vocalic portion or even just the vowel offset? The data here show good separability between the alveopalatal and the other sibilants at vowel offset, but additional perception work will need to be done to determine if listeners attend to this information and use it in perception for the purposes of sibilant discrimination.

4.7.3 Cue weighting in production and perception

The relationship between contrast and variation has implications for general understanding of cue weighting. Coefficients from linear discriminant analysis have often been used as the metric of cue strength/weight in production (Shultz et al., 2012; Garellek and White, 2015; Schertz et al., 2015; Kim and Clayards, 2019, among others). Shultz et al. (2012) examined multiple cues in production of English stops and found an inverse correlation between weight of VOT and weight of F0 (as determined by discriminant analysis coefficients). This indicates a type of trading relationship between the two cues: the more contrastively a speaker uses VOT, the less contrastively they use F0.

The hypothesis examined here is not that cue weight on the COG dimension correlates with cue weight on the F2 dimension, but rather that cue weight on the COG dimension correlates with variation on the F2 dimension. This is a related but non-identical hypothesis. However, an inverse relationship between the two relevant cues was also observed here, in line with the findings in Shultz et al. (2012). The results

Figure 4.14. Inverse relationship between weight of COG and weight of F2 across speakers.



are shown in Figure 4.14.[7] In this data, speakers who use COG more contrastively generally use F2 less contrastively and exhibited more variation in F2.

The results here build on the existing literature showing relationships between cue weights in production. As in the Shultz et al. (2012) English data, the relative strengths of the two cues are inversely correlated. The investigation here goes a step further by examining the relationship between the COG coefficients and F2 variation.[8] These results suggest it is possible to predict relative differences in extent of within-category variation on one dimension based on the cue weights of another dimension. If listeners know about this relationship, it could potentially aid in talker-specific adaption in perception.

---

[7]The coefficients are presented with their z-score values, as is done in Shultz et al. (2012); Schertz et al. (2015); Kim and Clayards (2019).

[8]Although it is intuitive that cue strength might correlate with within-category variation, this is not necessarily the case. It is possible to have a low cue weight due to high overlap with little within-category variation. For example, the values from two categories might be entirely overlapping yet have very little within-category variation.

Although correlations between relative cue weights have been observed in production, these relationships do not seem to be predictive of relative cue weights in perception. Multiple studies have examined the relationship between cue weights in production and cue weights in perception for the same individuals and have found no significant relationship (Schertz et al., 2015; Shultz et al., 2012; Kim and Clayards, 2019).Though they have found trends in the positive direction, potentially indicating a weak relationship between cue weight in production and cue weight in perception. Based on this, we would not expect the differences in COG contrast observed here to necessarily be predictive of how the speakers would use COG vs. F2 in a perception task.

4.7.4 Comparison with the between-language case studies

Both the within-language and between-language predictions of Contrast-Dependent Variation test the same general hypothesis that phonological contrast constrains possible phonetic variation such that variation emerges in the absence of contrast. The between-language prediction was tested with the Hindi/English and Polish/French case studies and the within-language prediction was tested in Mandarin. The within-language prediction examined in this chapter could also be tested in the other languages from the between-language case studies in future work.

For example, in English stops, I would predict that English speakers that have higher cue weight for the primary cue of VOT would also have higher amounts of within-category variation in secondary cues such as F0 or closure voicing. In fact, results from existing work on cue weighting in English stops (Shultz et al., 2012) (summarized in the previous section) already suggest that this might be the case. Shultz found an inverse relationship between weight of VOT and weight of F0. In the Mandarin data, I found a similar inverse relationship between weight of COG and weight of F2 and also a positive relationship between weight of COG and variation in

F2. If within-category variation is directly related to cue weight we would expect to see a similar correlation between weight of VOT and variation of F0 in English stops given the results here and the Shultz et al. (2012) findings.

4.8 Conclusion

In this chapter, I presented the results of a production study comparing F2 variation across Mandarin speakers. The hypothesis being tested is a prediction of the Contrast-Dependent Variation hypothesis examined in Chapters 2-3. In the case of the Mandarin sibilants, we expect relatively more variation along the F2 dimension for speakers that use COG more for contrast, which was observed in the data here. This result demonstrates a way in which patterns in extent of variation are predictable across speakers, an additional way in which phonetic variation is structured and not random. This has implications for Dispersion Theory, cue weighting, and perception.

CHAPTER 5

EVALUATING METRICS FOR DISPERSION AND SEPARABILITY

5.1 Introduction

This chapter[1] examines metrics used for quantifying dispersion between phonological categories, which are relevant methodologies for exploring the relationship between contrast and variation. I propose a new metric for calculating acoustic dispersion, which improves over the standard metric of mean-to-mean distance by incorporating within-category variance information directly into the distance measurement.

The Contrast-Dependent Variation hypothesis tested in this dissertation is a revision of Lindblom's (1986) hypothesis which is based in Dispersion Theory (DT). DT makes predictions about the language-specific phonetic realization of phonemes (variation, spread, etc.) and about which phonological inventories should be typologically common (see Chapter 1 for a detailed review). In this chapter, I examine the large-scale typological predictions of DT and their application to consonant inventories, using modeled data from Schwartz et al. (2012).

Previous work has found that the most common stop inventory is not the most acoustically dispersed unless pharyngeals and epiglottals are excluded from the dispersion calculation (Schwartz et al., 2012). Although the new metric proposed here does not recover DT predictions for stop inventories, it changes results, showing that dispersion results depend on metric choice. The metric can be used in any acoustic

[1]Portions of this chapter have been published previously in Hauser (2017).

space to include information about within-category variance when calculating dispersion.

5.2 Background

Dispersion Theory (Liljencrants and Lindblom, 1972; Schwartz et al., 1997) has been used to account for typological trends in vowel inventories cross-linguistically. Dispersion[2] is typically calculated using triangle area between three mean or prototypical vowel points in the acoustic space of formant frequencies. Within-category variance has been proposed to be a factor in Dispersion Theory (Lindblom, 1986) and data exists showing it affects perception (Pisoni and Tash, 1974; McMurray et al., 2002; Clarke and Luce, 2005; Clayards et al., 2008), yet conventional dispersion metrics do not take within-category variance into account. A new metric for calculating dispersion is proposed here which incorporates within-category variance. As a test case for these analyses, I examine place of articulation dispersion in stop inventories.

Schwartz et al. (2012) used modeled vocal tract articulatory-acoustic data to evaluate acoustic dispersion in stop inventories. They used mean-to-mean distance between acoustic categories in three-dimensional formant space $< F1, F2, F3 >$ to calculate dispersion of stop place of articulation (POA).[3] With this metric, Schwartz et al. found that the typologically common /bilabial coronal velar/[4] configuration is not the most dispersed three-stop inventory in the acoustic space of formant onset frequencies (which are taken to be the primary perceptual cue for POA, see §5.2.2 for

[2] *Dispersion Theory* is used to reference the theory posited by Liljencrants and Lindblom (1972) and other subsequent work which makes predictions about typological frequency of certain inventories. This is distinct from *dispersion* itself which refers to a measure capturing how "spread out" points or categories are and is not necessarily tied to the predictions of *Dispersion Theory*.

[3] They discuss the use of mean-to-mean distance, though no quantitative metric is provided.

[4] Effectively all languages with three-stop inventories (or stop inventories which use three places of articulation plus voicing contrasts) have a /bilabial coronal velar/ configuration. 334 out of 336 languages with three-stop inventories in P-base (Mielke, 2008) are /bilabial coronal velar/.

further discussion). Instead, the optimally dispersed configuration is the unattested (not observed in any languages of the world) /coronal velar epi-pharyngeal/ inventory. Schwartz et al. (2012) did not find that the most dispersed inventory in the acoustic space of the first three formants is the most typologically common. This is inconsistent with DT, which predicts that the most common inventories will be acoustically dispersed to aid in perceptual distinction.

Schwartz et al. (2012) recover the DT result by proposing a restriction of the phonetic space according to an evolutionary and articulatory Frame Content model (MacNeilage, 1998) and their own Perception for Action Control Theory to exclude pharyngeals and epiglottals. Perception for Action Control Theory claims that the gestural content of speech constrains the perceptual representation. They consider speech sounds to be "bundles" of articulatory and perceptual features, integrating previously competing perceptual and motor theories.

This is combined with Frame Content Theory (MacNeilage, 1998), an evolutionary account of the emergence of language which hypothesizes that proto-consonants emerged from the high mandible cycles of jaw closure and proto-vowels from the low cycles. The theory also applies to child language acquisition as the child begins babbling with similar mandible cycles. With progressive exploration of the vocal tract and gradual enlargement of the stop space from a neutral vocal tract configuration, /b d g/ emerges as the optimal stop system in terms of acoustic dispersion in the space allowed by Frame-Content style exploration. Schwartz et al. model this type of exploration using the same vocal tract model which generated the stop space. Pharyngeals and epiglottals are excluded because their articulation involves a downward movement of the mandible and is therefore not predicted to be a proto-stop by Frame Content Theory.

However, while this argument provides an explanation for why /b d g/ may have been evolutionarily prior to a system with pharyngeals and epiglottals, it does not

explain why this configuration is so common presently. Pharyngeal and epiglottal stops are attested and should not be harder to produce than other stops (Edmondson et al., 2005), so we might expect that they would have become more ubiquitous over time given that they create a dispersed triangle with /d/ and /g/.

In light of the Schwartz et al. results, there are three hypotheses which could explain why the most dispersed stop inventory is not typologically common: (1) Dispersion Theory does not apply to consonants, (2) the phonetic space in which dispersion was considered is not appropriate for stop POA, or (3) the metric by which dispersion was calculated is not the most relevant for the data. The second hypothesis was explored by Schwartz et al. (2012). In a phonetic space which does not include pharyngeal or epiglottal stops, the /bilabial coronal velar/ inventory was the most dispersed. In that analysis, Dispersion Theory did not apply to stop inventories without manipulating the phonetic space when the conventional mean-to-mean distance metric was used. This chapter tests the third hypothesis: a different metric for dispersion is needed to capture distributional information about the categories instead of only their central tendencies.

5.2.1 Relevance of variance information

The acoustic categories of stop consonants examined in this chapter do not have homogeneous distribution shapes or variances, and are therefore not well represented only by their means. Collapsing acoustic categories to their means results in loss of important distributional information. While this dissertation does not directly examine perception, DT was originally argued to have a perceptual explanation (Liljencrants and Lindblom, 1972); acoustic categories should be dispersed to aid in perceptual distinction. Therefore, the methods used to test the predictions of DT should be informed by perceptual data as best as possible.

Including distributional information in a dispersion metric instead of only category means is perceptually relevant because within-category distributional information has been shown to affect human speech perception. Clayards et al. (2008) provides experimental data which shows that perceptual uncertainty increases with within-category variation. There is also evidence that listeners are sensitive to within-category distributional differences based on eye movements (McMurray et al., 2002), and boundary marking between categories (Clarke and Luce, 2005). Given these findings, within-category variation should be considered when calculating acoustic dispersion.

5.2.2 Formants as space for POA

Using the first three formants measured at the onset of the vowel is a reasonable phonetic space for considering dispersion of stop categories. Stop articulation is complex and various acoustic measures could potentially be used to classify stops including burst spectrum and voice onset time, but there is experimental and computational evidence suggesting the primacy of formant transitions as cues to place of articulation.

Stevens and Blumstein (1978) argue that the onset spectrum of the CV syllable provides a primary and invariant cue to POA. In Walley and Carrell (1983), adults and children were asked to identify stops where the burst spectra and formant transitions specified different places of articulation. Both adults and children consistently used the formant transitions for place of articulation identification. This provides perceptual evidence that formant transitions are primary cues to place of articulation in stops. In computational work, Sussman et al. (1991) found that locus equations effectively categorize stops by place of articulation. Locus equations use a regression line fit between F2 onset and F2 vowel values following stop consonants; using a model which employs locus equations can categorize stops by place with little error.

This provides computational evidence for the importance of formant transitions in the categorization of stops by place of articulation.

While these results provide evidence for the importance of formant transitions, the values used in Schwartz et al. (2012) (and in the calculations here) are not transition values, but onset formant values. These values do not necessarily convey information about how the formant value changes over time. Therefore, they are not directly analogous to the results discussed above. However, I do follow Schwartz et al. (2012) in assuming these onset formant measurements to be an appropriate (though perhaps non-optimal) space for considering dispersion of stop POA.

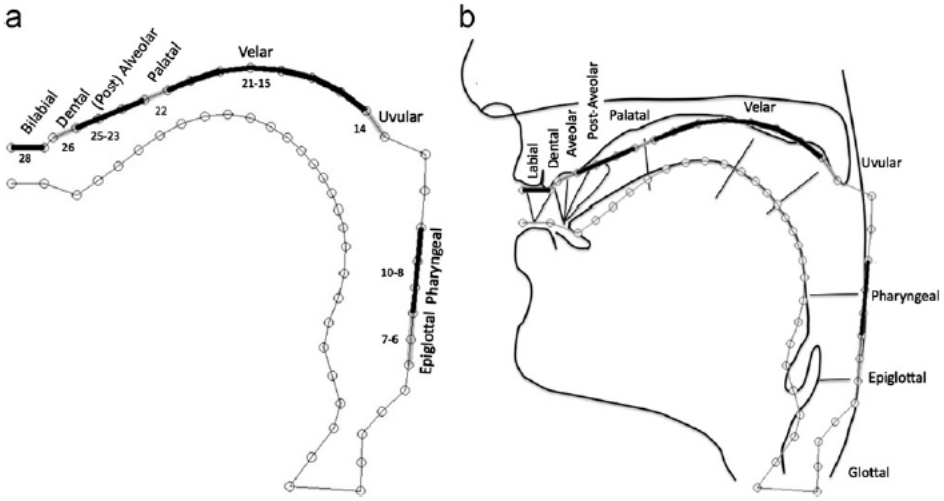
5.3 Methods

The data analyzed here are the same data generated by a vocal tract articulatory-acoustic model in Schwartz et al. (2012).[5] The vocal tract model is based on the vocal tract model originally developed by Maeda (1990) which was designed from drawings of over 500 hand drawn mid-sagittal contours obtained from the reading of ten French sentences. The vocal tract was divided into a standard set of sections from the lips to the glottis which defined possible places of articulation, shown in Figure 5.1. The possible closures were distributed along the grid in Figure 5.1(a). Figure 5.1(b) shows how the grid corresponds to places of articulation as reported by Ladefoged and Maddieson (1996).

50,000 stop tokens were generated with possible occlusions at all places along the vocal tract between lips and glottis in three vowel contexts _[i a u]. Double articulations and non-anatomically possible articulations were excluded. The tokens were randomly sampled from the grid of the vocal tract in 5.1(a).

[5]Many thanks to Schwartz et al. for sharing this data and providing helpful commentary.

Figure 5.1. Vocal tract model places of articulation from the model in Schwartz et al. (2012)

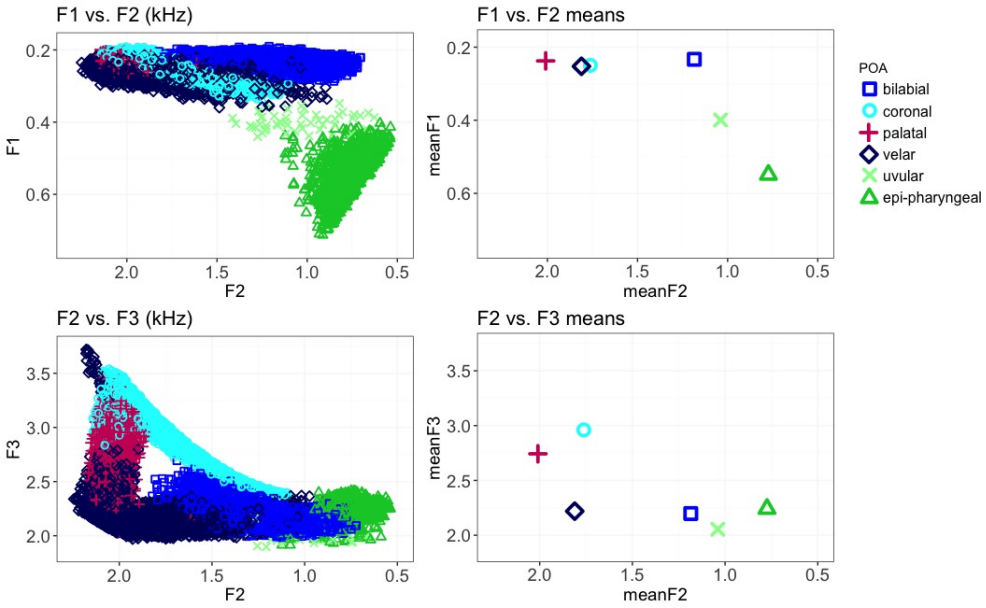


The stop tokens were grouped according to the following places of articulation (POA) on the vocal tract: bilabial, coronal, palatal, velar, uvular, epi-pharyngeal.[6] After the stop burst, when the vocal tract is no longer occluded, formant structure (resonances of the vocal tract) appears due to the production of the vowel. The first three formants (F1, F2, F3) were measured at the beginning of the transition to the vowel where the formant structure first appears after the occlusion. These formant measurements served as the input for the calculations presented in the following sections.

Figure 1[7] shows the data plotted in $< F1, F2 >$ (top panel) and $< F2, F3 >$ (lower panel) space. The plots on the left display all of the data from the model, coded by color and shape for POA. The plots on the right show only the means at each POA. It is evident from this visualization that the variances and shapes of the

[6]The model produced separate categories for alveolar and dental, but the categories are collapsed into one category coronal due to the rarity of contrastive dental place cross-linguistically. Only 3% of languages contrast dental and alveolar place (Mielke, 2008). Epiglottal and pharyngeal stops are also merged for similar reasons (Esling, 2003).

[7]All calculations were done in R (R Core Team, 2013) with plotting using the ggplot2 package (Wickham, 2009).



stop categories in this acoustic space are not homogeneous. Reducing these categories to a single mean point obscures this distributional information, which is relevant for evaluating dispersion.

5.4 Mean-to-mean acoustic distance

In this section, dispersion calculations are presented which utilize triangle area and a definition of mean-to-mean distance based on Euclidean distance. Triangle area in formant space is the conventional way of calculating dispersion in vowel spaces (Andruski et al., 1999; Jacewicz et al., 2007). These results replicate the findings of Schwartz et al. (2012) who used mean-to-mean distance in three-dimensional formant space but did not provide a quantitative definition, so it can be assumed the standard triangle area approach was used.

The data generated by the model include a cluster of points at each place of articulation in three-dimensional $< F1, F2, F3 >$ space. The mean of each cluster was calculated and used to measure dispersion between every possible combination of three place of articulation categories. Distance between two mean points in the three-dimensional formant space was calculated using the equation in Figure 5.2, and

Figure 5.2. Equation: Distance between two mean points (i, j) in $< F1, F2, F3 >$ space

$$d_{ij} = \sqrt{(F1_j - F1_i)^2 + (F2_j - F2_i)^2 + (F3_j - F3_i)^2}$$

Figure 5.3. Equation: Area of a triangle as a dispersion measure

$$A = \sqrt{s(s - d_{ij})(s - d_{jk})(s - d_{ik})} \qquad \text{where} \qquad s = (d_{ij} + d_{jk} + d_{ik})/2$$

dispersion was measured using the area of the triangle made by three means as in the equation in Figure 5.3. By definition, the larger the area of the triangle, the more dispersed the three points are.

Dispersion was calculated for all possible combinations of three places of articulation (20 total). Table 5.1 provides dispersion measures for the five most dispersed three-stop inventories in the three dimensional $< F1, F2, F3 >$ space in kHz^2 and ERB^2. Use of both Hz and a perceptual unit to calculate dispersion is in accordance with previous work on Dispersion Theory (Lindblom, 1986). Transformation into either ERB or Bark slightly changes a few of the rankings, but the /bilabial coronal velar/ inventory consistently ranks below several unattested inventories regardless of unit of measure for formant frequencies in these calculations.

5.4.1 Interim discussion: Mean-to-mean acoustic distance

The measures in Table 5.1 support the conclusion of Schwartz et al. (2012) that the /coronal velar epi-pharyngeal/ inventory is the most dispersed in this acoustic space

Table 5.1. Mean-to-mean distance dispersion results (ranked according to kHz^2 results)

	POA1	POA2	POA3	Dispersion (kHz^2)	Dispersion (ERB^2)
1	coronal	velar	epi-pharyngeal	0.40	9.6
2	coronal	velar	uvular	0.30	6.2
3	palatal	velar	epi-pharyngeal	0.29	7.1
4	bilabial	coronal	epi-pharyngeal	0.23	11
5	**bilabial**	**coronal**	**velar**	**0.23**	**4.2**
	...				

of $< F1, F2, F3 >$. The typologically common inventory, /bilabial coronal velar/, is the fifth most dispersed inventory. All inventories which are better dispersed than /bilabial coronal velar/ are unattested. Based on this mean-to-mean distance definition of dispersion in this acoustic space, Dispersion Theory does not make accurate predictions regarding typological frequency.

But mean-to-mean distance may not be a good metric for understanding the dispersion of stop categories within this space, as discussed in §5.2. While it does capture the central tendency of each category, mean-to-mean distance ignores the distribution of all the tokens, reducing each category to a single data point. It could be the case that the /coronal velar epi-pharyngeal/ inventory, while dispersed under the mean-to-mean distance metric, is non-optimal because high amounts of within-category variance reduce dispersion. Given consistent mean-to-mean distance between a set of distributions, a set with tighter variances is more dispersed than a set of distributions with greater variances and therefore more overlap. Incorporating covariance into a dispersion metric better captures these distributional aspects of dispersion.

5.5 Incorporating variance: Jeffries-Matusita distance

To incorporate within-category distribution information, I propose a dispersion metric which incorporates the Jeffries-Matusita (JM) distance (Bruzzone et al., 1995; Kobayashi and Thomas, 1967; Jolad et al., 2012). This distance metric incorporates covariance, the multidimensional analog of variance. The JM distance is a transformation of the Bhattacharrya distance, which is often used as a class separability measure in feature selection and pattern recognition literature (Choi and Lee, 2003). Although these are standard measures in other literatures, neither distance metric has been applied to calculate acoustic dispersion in speech sound inventories. The JM distance transforms the Bhattacharrya distance into a fixed range $[0,\sqrt{2}]$ instead of an infinite range. Other distance metrics incorporating covariance were also used

Figure 5.4. Equation: Jeffries-Matusita Distance (D_{JM}) as a function of the Bhattacharya Distance (D_B) between two Gaussian distributions F, G with probability density functions f, g

$$D_{JM}(F,G) = \sqrt{2(1 - exp(-D_B(F,G)))} \quad \text{where} \quad D_B(F,G) = \int_x \sqrt{f(x)g(x)}dx$$

and obtained similar results, which suggests these results are due to the incorporation of covariance generally, not the JM distance specifically.

The general equation for the JM distance (Bruzzone et al., 1995) is given in the equation in Figure 5.4. The JM distance, unlike mean-to-mean distance, applies to distributions[8] rather than single points. Using the JM distance instead of mean-to-mean distance as the base of the dispersion metric incorporates the geometric fact that dispersed inventories have categories which have large between-category variation relative to the amount of within-category variation, and therefore less category overlap. Dispersion is still calculated based on triangle area as in Equation 5.3 in $< F1, F2, F3 >$ space, but using JM distance measures instead of mean-to-mean Euclidean distance measures.

The results from the JM distance calculations are given in Table 5.2, which shows a selection of the 16 most dispersed three-stop inventories.[9] These results are different from the mean-to-mean distance calculations; the most dispersed inventory is now /bilabial coronal epi-pharyngeal/, an inventory which is also unattested. There are 15 inventories which are more dispersed than the typologically common /bilabial coronal velar/ inventory.

---

[8]This equation assumes Gaussian distributions. This is a standard assumption made for ease of computation. Restructuring and transforming the data to better approximate Gaussian distributions does not change results.

[9]These dispersion measures are unitless because the JM distance transforms the distances into the fixed range $[0, \sqrt{2}]$ which are then used to calculate triangle area, so the resulting dispersion measures will always lie within the same fixed range regardless of the original unit of the formant measurements.

Table 5.2. JM distance dispersion results

	POA1	POA2	POA3	Dispersion (from Hz)	Dispersion (from ERB)
1	bilabial	coronal	epi-pharyngeal	0.865	0.865
2	bilabial	palatal	epi-pharyngeal	0.864	0.850
3	bilabial	palatal	uvular	0.864	0.850
⋮					
14	coronal	palatal	uvular	0.782	0.789
15	bilabial	coronal	palatal	0.781	0.776
16	**bilabial**	**coronal**	**velar**	**0.774**	**0.741**

5.5.1 Interim discussion: JM distance

The use of the Jeffries-Matusita distance changed results, providing an example of how choice of metric is crucial in work on dispersion of inventories. However, using the Jeffries-Matusita distance for stop inventories did not improve results in favor of the predictions of Dispersion Theory relative to the use of mean-to-mean distance despite including covariance, additional perceptually relevant information. The stop system which should be optimal given its typological frequency, /bilabial coronal velar/, is not optimal in the results of this study in the $< F1, F2, F3 >$ space with either metric for calculating dispersion. In the results here, the epi-pharyngeal place is always in the most dispersed inventory, likely because of the structure of the distributions in the three dimensional formant space. Epi-pharyngeal consonants are distinct from all the other places of articulation along the F1 dimension (this is visible in Figure 1).

One major difference in the JM distance results relative to the mean-to-mean distance results is the lack of the velar place in any of the top five most dispersed inventories. In the most dispersed inventory, bilabial replaced velar when the analysis was done with JM distance. This is due to the high amounts of within-category variation in the formants following the velar stops. This can be viewed in Figure 5.3. The mean of the velar category is not representative of the entire distribution and the high amounts of within-category variation result in higher measures of the JM distance.

Table 5.3. JM distance dispersion results: $< F2, F3 >$ space

	POA1	POA2	POA3	Dispersion (from Hz)	Dispersion (from ERB)
1	coronal	velar	epi-pharyngeal	0.836	0.840
2	coronal	palatal	epi-pharyngeal	0.774	0.783
3	coronal	palatal	uvular	0.765	0.777
⋮					
9	palatal	velar	epi-pharyngeal	0.741	0.764
10	bilabial	coronal	epi-pharyngeal	0.737	0.735
11	**bilabial**	**coronal**	**velar**	**0.698**	**0.665**

A conceivable change to the space would be to down-weight or exclude the F1 dimension entirely, on the premise that it must not be relevant for consonant perception given the rarity of epi-pharyngeal stops cross-linguistically. However, considering only an acoustic space of $< F2, F3 >$ does not automatically make /bilabial coronal velar/ optimally dispersed. The /coronal velar epi-pharyngeal/ inventory is still the most dispersed three-stop inventory even when the F1 dimension is excluded entirely. The dispersion results in the $< F2, F3 >$ space are provided in Table 5.3. The predicted /bilabial coronal velar/ inventory is still not the most dispersed inventory, and all the inventories which are better dispersed are either typologically rare or unattested.

Following Lindblom's (1986) revision of Dispersion Theory, it could be the case that the /bilabial coronal velar/ inventory is not maximally dispersed, but is *sufficiently* dispersed. The /coronal velar epi-pharyngeal/ inventory could be the maximally dispersed inventory but the dispersion of /bilabial coronal velar/ is sufficient and therefore optimal. If this were the case, inventories with similar dispersion measures to /bilabial coronal velar/ should also be typologically common but effectively all other three-stop inventories are unattested. Drawing a distinction between sufficient and maximal dispersion does not immediately explain why /bilabial coronal velar/ is typologically common but not dispersed.

5.6 Conclusion

These results showcase the methodological importance of choosing the appropriate metric for quantifying acoustic dispersion. I argue that incorporating within-category variance should be a primary consideration in work on acoustic dispersion. The most dispersed inventory in $< F1, F2, F3 >$ and $< F2, F3 >$ space is not the most typologically common inventory, but is instead an inventory which is unattested. Given that the formant space is the appropriate phonetic space for stop place of articulation (see §5.2) and the metric is suitable, these results suggest that the typological predictions of Dispersion Theory do not apply to consonants as they apply to vowels. The predictions of Dispersion Theory are not recovered by using a metric which incorporates distributional information. If Dispersion Theory does apply to stop inventories, it must be the case that the phonetic space is altered in some way (e.g. Schwartz et al. (2012)).

The new dispersion metric proposed here builds on the mean-to-mean distance measure for dispersion by incorporating important perceptually relevant information about category distributions. The metric is not specific to stop inventories and can be used to calculate dispersion in any acoustic space. The results show that incorporating within-category variance information into a dispersion metric does affect the outcome of dispersion results. This is important to consider in work on Dispersion Theory, regardless of the acoustic space under consideration.

CHAPTER 6

CONCLUSION

6.1 Summary

This dissertation has investigated the effect of phonological contrast on phonetic variation in multiple case studies. The hypotheses here broadly propose that variation emerges in the absence of contrast, resulting in predictable language- and speaker-specific differences in extent of phonetic variation. This roughly follows a prediction of Dispersion Theory (Liljencrants and Lindblom, 1972; Lindblom, 1986), which states that phonological systems with more contrasts should exhibit less phonetic variation relative to systems with fewer contrasts. However, the results here have supported a revision of this prediction which refers to specific phonetic dimensions rather than phonological systems/inventories.

The first test case presented in Chapter 2 examined stops consonants in Hindi and English. Hindi has four stop contrasts at each place of articulation, while English has two. Contrast-Dependent Variation predicts less variation in Hindi, but only along the particular phonetic dimensions which realize additional phonological contrasts relative to English. In the results of the experimental task, both languages showed similar amounts of group level within-speaker variation, but English speakers exhibited more group level within- and between-speaker variation in closure voicing. This is in accordance with Contrast-Dependent Variation as the voicing dimension is employed as the primary cue to phonological contrasts in Hindi but not in English.

The second test case presented in Chapter 3 examined sibilant fricatives in Polish and French. Polish has three voiceless sibilants which contrast in place of articulation,

while French has two. Contrast-Dependent Variation predicts no difference in amount of within-category variation in COG or F2 in these two languages. This is because both phonetic dimensions serve as primary cues to phonological contrasts in the context elicited in the experiment. This differs from the predictions of Lindblom (1986). A direct implementation of that hypothesis would likely consider the sibilant inventory as the relevant "system" and predict more variation in French because Polish has more sibilant phonemes. This case study provides a concrete example of how Contrast-Dependent Variation differs from Lindblom (1986) by considering phonetic dimensions rather than subsets of phonological inventories.

The final test case presented in Chapter 4 is a within-language case study examining differences in extent of variation in production of sibilant fricatives across speakers in Mandarin. The hypothesis is an extension of the between-language hypothesis which predicts relative differences in extent of variation between speakers who use different phonetic dimensions to realize the same phonological contrast.

The last chapter explores methods for quantifying contrast through metrics of dispersion and separability. A new metric for calculating acoustic dispersion is proposed and is compared to the traditional mean-to-mean distance using modeled stop inventory data. The results show that choice of metric changes results, although the new metric does not recover the cross-linguistic typological predictions of Dispersion Theory for stop inventories.

6.2 Contributions and implications

6.2.1 Dispersion Theory

This dissertation contributes to the literature on Dispersion Theory (DT) by explicitly testing one of the hypotheses in Lindblom (1986): that phonemes in larger inventories should show relatively less within-category variation than phonemes in smaller inventories. In its original formulation, this hypothesis only refers to vowel

inventories, yet it is sometimes assumed to be general. I have explicitly tested this hypothesis and proposed a revision (Contrast-Dependent Variation) which is more explicit and domain general.

The work here also extends the predictions of DT to consonant inventories, following Boersma and Hamann (2008); Schwartz et al. (2012). The previous work has focused solely on dispersion of the central tendencies of categories in acoustic space. This dissertation considers the related and often assumed prediction of DT, that within-category variation should correlate with size of inventory. My proposal refines this, making the hypothesis about phonetic dimensions instead of phonemes. I argue that the prediction must be operationalized over phonetic dimensions rather than (potentially ad hoc) subsets of phonological inventories. Under this reformulation, we do expect relatively more variation in smaller inventories but only along the specific phonetic dimensions that realize additional contrasts. Because this hypothesis is domain general, it can potentially be applied to vowel inventories as well in future work.

6.2.2 Structure in phonetic variation

There is a large body of work showing that variation in phonetic realization is structured and not entirely random (see Chapter 1 for a review). The results in this dissertation build on those findings by showing additional ways in which phonetic variation in systematic. Previous work mostly focuses on explaining fluctuations in acoustic values according to various non-contrastive conditioning factors. My findings not only show structure in phonetic variation, they additionally show systematicity in *extent* of variation across languages and speakers.

6.3 Remaining questions and future work

In order to further refine these hypotheses, additional test cases are necessary. There are two particular stop cases which seem to be likely counterexamples for the between-language prediction of Contrast-Dependent Variation. As I have formulated the hypothesis here, the only factors contributing to relative differences in within-category variation is the presence of phonological contrast on particular phonetic dimensions in the language. This may be too simplistic to generalize to all cases and more work needs to be done to further clarify the hypothesis.

Given the findings that patterns of variation are language- and speaker-specific, we might also ask how malleable these patterns are across different speaking contexts. Specifically, do speakers adjust the extent of variation in production in the direction of recently heard speech? These questions could be addressed by employing the phonetic imitation paradigm which uses unconscious human imitative tendencies to examine the link between speech perception and production. Foundational work using this methodology has shown that speakers do not imitate phonetic values that approach a phonological category boundary (Nielsen, 2011). Before examining imitation of variation, additional testing needs to be done to remove a potential confound of hypoarticulation, which is present in previous work. This will clarify the effect before generalizing to different languages and examining imitation of variation.

The findings in this dissertation also have implications for L2 acquisition and non-native speech production. Recent work on non-native speech production has shown that extent of phonetic variability is specific to the L1/L2 pairing (Vaughn et al., 2018). It is not the case that non-native speakers are always more variable than native speakers. Rather, the relative differences between variation in native and non-native speech seem to be specific to the phonetic dimension under consideration and the L1/L2 pairing.

Vaughn et al. (2018) examines patterns of variability in vowels and stops produced by native Japanese speakers and L2 Japanese speakers with L1s English and Mandarin. They did not observe significant differences in the amount of variability in Japanese vowel formants between L2 Japanese L1 Mandarin speakers or L2 Japanese L1 English speakers relative to L1 Japanese speakers. However, there were non-significant trends of less variation for L1 Mandarin speakers and more variation for L1 English speakers. In stop productions, L2 Japanese L1 Mandarin speakers showed more variability in VOT of voiceless stops relative to L1 Japanese speakers, but less variability in VOT of voiced stops. L2 Japanese L1 English speakers did not differ from the L1 Japanese speakers. They also found no differences in means or variability in non-native speech between learners with different amounts of instruction and exposure to Japanese.

These findings suggest that additional properties of the L1 (other than mean values) may affect L2 production. Vaughn et al. (2018) offer a potential explanation for their findings by invoking the Dispersion Theory hypothesis from Lindblom (1986) examined here. If native speakers produce different amounts of within-category variation due to different inventory sizes, it could be the case that native speakers transfer that phonetic variability into their L2 production. This would result in the differences observed by Vaughn et al. (2018). However, more work would need to be done to clarify this hypothesis.

Vaughn et al. suggest that differences in L1 variability may play a role in determining variability in L2 production. The results in this dissertation are relevant to these conclusions as they provide evidence for language-specific L1 variability patterns. The Vaughn et al. (2018) results could be paired with Contrast-Dependent Variation for the purposes of further developing theories of phonetic transfer in L2 speech to additionally account for phonetic variability. More work will need to be done to determine whether incorporating variation into existing theories of phonetic

transfer is warranted. If native language variability patterns transfer in L2 acquisition, the hypotheses proposed here make predictions not only about language-specific variability, but also about variability patterns in non-native speech.

APPENDIX A

HINDI AND ENGLISH STOPS

A.1 Full output of regression models

Table A.1. English full model with word as random intercept and logit link for beta regression. Call: voicing percent ∼ voicing + V × speaker + place × V + place × speaker + closure duration + block + (1|word). Model intercept is speaker e02 block 1 voiced coronal /a/ context.

	Estimate (se)	z	p
(Intercept)	2.45 (0.26)	9.38	$< 0.001^{***}$
voicing-voiceless	-1.57 (0.08)	-18.52	$< 0.001^{***}$
V-/i/	0.63 (0.23)	2.70	0.007^{**}
V-/u/	-0.42 (0.25)	-1.70	0.089
speaker-e03	0.55 (0.23)	2.37	0.018^{*}
speaker-e04	-0.68 (0.23)	-2.97	0.003^{**}
speaker-e06	-1.05 (0.23)	-4.59	$< 0.001^{***}$
speaker-e07	-1.95 (0.24)	-8.28	$< 0.001^{***}$
speaker-e09	-1.92 (0.24)	-8.14	$< 0.001^{***}$
place-labial	-0.51 (0.24)	-2.15	0.032^{*}
place-velar	-0.22 (0.23)	-0.95	0.344
closure duration	-10.08 (1.23)	-8.18	$< 0.001^{***}$
block 2	0.06 (0.08)	0.72	0.471
block 3	-0.05 (0.08)	-0.56	0.573
block 4	0.05 (0.08)	0.63	0.530
V-/i/:speaker-e03	-0.45 (0.25)	-1.79	0.074
V-/u/:speaker-e03	0.82 (0.26)	3.22	0.001^{**}
V-/i/:speaker-e04	-0.06 (0.25)	-0.23	0.820
V-/u/:speaker-e04	1.38 (0.26)	5.36	$< 0.001^{***}$
V-/i/:speaker-e06	-0.54 (0.26)	-2.08	0.037^{*}
V-/u/:speaker-e06	0.50 (0.26)	1.90	0.058
V-/i/:speaker-e07	0.08 (0.25)	0.32	0.750
V-/u/:speaker-e07	1.00 (0.26)	3.81	$< 0.001^{***}$
V-/i/:speaker-e09	-0.11 (0.26)	-0.42	0.676
V-/u/:speaker-e09	0.84 (0.27)	3.15	0.002^{**}
V-/i/:place-labial	-0.05 (0.22)	-0.22	0.829
V-/u/:place-labial	0.09 (0.24)	0.36	0.719
V-/i/:place-velar	0.26 (0.22)	1.18	0.239
V-/u/:place-velar	0.28 (0.23)	1.22	0.221
speaker-e03:place-labial	0.51 (0.26)	1.99	0.046^{*}
speaker-e04:place-labial	0.13 (0.26)	0.49	0.627
speaker-e06:place-labial	0.37 (0.27)	1.38	0.167
speaker-e07:place-labial	0.70 (0.26)	2.67	0.008^{**}
speaker-e09:place-labial	0.51 (0.27)	1.92	0.055
speaker-e03:place-velar	-0.25 (0.25)	-1.00	0.319
speaker-e04:place-velar	-0.70 (0.25)	-2.77	0.006^{**}
speaker-e06:place-velar	-0.52 (0.26)	-1.97	0.049^{*}
speaker-e07:place-velar	0.01 (0.26)	0.04	0.970
speaker-e09:place-velar	-0.12 (0.26)	-0.44	0.662
Num. obs	1561		
Num. groups: word	61		
Var. word (intercept)	0.034		

Table A.2. Hindi full model with word as random intercept and logit link for beta regression. Call: voicing percent ∼ voicing + V × speaker + place × V + place × speaker + closure duration + block + (1|word). Model intercept in speaker h09 block 1 voiced coronal /a/ context.

	Estimate (se)	z	p
(Intercept)	3.83 (0.22)	17.68	$< 0.001^{***}$
voicing-voiceless	-4.94 (0.09)	-55.55	$< 0.001^{***}$
V-/i/	0.08 (0.19)	0.40	0.692
V-/u/	0.11 (0.19)	0.57	0.565
speaker-h11	0.29 (0.20)	1.45	0.148
speaker-h13	-0.01 (0.20)	-0.06	0.953
speaker-h14	0.00 (0.20)	-0.02	0.980
speaker-h15	0.09 (0.21)	0.43	0.667
speaker-h16	-0.06 (0.21)	-0.26	0.792
place-labial	0.16 (0.23)	0.69	0.487
place-velar	0.06 (0.19)	0.29	0.773
closure duration	-2.84 (0.68)	-4.20	$< 0.001^{***}$
block 2	-0.14 (0.08)	-1.75	0.079^{*}
block 3	-0.20 (0.08)	-2.40	0.016^{*}
block 4	-0.21 (0.08)	-2.53	0.012^{*}
V-/i/:speaker-h11	-0.03 (0.25)	-0.12	0.903
V-/u/:speaker-h11	-0.27 (0.24)	-1.11	0.265
V-/i/:speaker-h13	0.06 (0.25)	0.24	0.811
V-/u/:speaker-h13	0.11 (0.25)	0.44	0.659
V-/i/:speaker-h14	-0.12 (0.24)	-0.48	0.632
V-/u/:speaker-h14	0.07 (0.24)	0.27	0.786
V-/i/:speaker-h15	-0.07 (0.24)	-0.29	0.773
V-/u/:speaker-h15	0.03 (0.24)	0.11	0.910
V-/i/:speaker-h16	0.29 (0.25)	1.20	0.231
V-/u/:speaker-h16	0.07 (0.24)	0.30	0.763
V-/i/:place-labial	-0.17 (0.19)	-0.89	0.373
V-/u/:place-labial	-0.05 (0.18)	-0.27	0.786
V-/i/:place-velar	0.13 (0.16)	0.81	0.420
V-/u/:place-velar	-0.06 (0.16)	-0.40	0.691
speaker-h11:place-labial	-0.17 (0.26)	-0.63	0.526
speaker-h13:place-labial	-0.16 (0.28)	-0.57	0.572
speaker-h14:place-labial	-0.26 (0.28)	-0.95	0.341
speaker-h15:place-labial	0.00 (0.27)	0.01	0.996
speaker-h16:place-labial	0.01 (0.27)	0.05	0.961
speaker-h11:place-velar	0.09 (0.23)	0.40	0.692
speaker-h13:place-velar	-0.25 (0.24)	-1.06	0.290
speaker-h14:place-velar	-0.19 (0.23)	-0.84	0.404
speaker-h15:place-velar	-0.20 (0.22)	-0.90	0.366
speaker-h16:place-velar	-0.29 (0.23)	-1.28	0.202
Num. obs	1455		
Num. groups: word	69		
Var. word (intercept)	1.34e-09		

A.2 Forced alignment in Hindi

Both the English and Hindi data in Chapter 2 were forced aligned using the Montreal Forced Aligner (MFA McAuliffe et al., 2017). MFA is an open-source software which can be downloaded at https://montreal-forced-aligner.readthedocs.io/. The aligner requires an acoustic model, a transcript, and a pronunciation dictionary as input. The output is a set of textgrids which correspond to the audio files and have two tiers: one for word boundaries and one for phone boundaries.

The English data were aligned using the pretrained acoustic model and pronunciation dictionary for English (non-word stimuli were manually added to the pronunciation dictionary), both of which are available at the above site. The transcript was automatically generated from the list of experimental stimuli.

There was no pretrained acoustic model for Hindi at the time of analysis. To align the Hindi data, I trained an acoustic model on the data from the experiment which was then used to perform the alignment on the same data. This was done by using the train_and_align function of the MFA. To train a new acoustic model, the MFA requires a transcript and a pronunciation dictionary, or a phonemic transcript. The Hindi phonemic transcript was easily generated from the stimuli list and the data were able to be aligned with the transcript and raw audio files. The output of train_and_align is both the aligned textgrids described above and a new acoustic model which can be used to align future data. The acoustic model also includes an automatically generated dictionary of all the words in the transcripts.

The acoustic model which I trained on the Hindi data is available for download in the public archive for this experiment. Since it was only trained on this experimental data, the corresponding dictionary only includes words used in the experiment. The dictionary would need to be updated for more general use.

A.3 Stop voicing analysis with speaker as a random effect

As discussed in §2.5.3.3, there are multiple reasons to prefer an analysis which uses speaker as a fixed effect for this particular data. However, the results do not meaningfully change if speaker is instead included as a random effect in the regression models.

The full models are given in Tables A.4–A.3. There are few differences when speaker is included as a random effect. In the English model, there is still a significant difference between the velar place and the coronal intercept. There are also significant vowel differences for both high vowels relative to the low vowel /a/ instead of just /i/. The effect of closure duration is still significant. The Hindi model does not show any differences in the remaining factors when speaker is included as a random effect.

In this analysis, the best fit model for English is the full model. The best fit model for Hindi includes the factors of voicing and closure duration. This is the same model which is selected as the best fit model when speaker is included as a fixed effect (model output given in Table 2.10). Speaker as a predictor is included in the best fit models for English and not included in the best fit models for Hindi regardless of whether speaker is specified as a random or fixed effect.

Table A.3. Hindi full model with word and speaker as random intercepts and logit link for beta regression. Call: voicing percent ∼ voicing + place × V + closure duration + block + (1 | word) + (1 | speaker)

	Estimate	Std. Error	z	p
(Intercept)	3.811	0.132	28.863	0.000
voicing-voiceless	-4.928	0.089	-55.372	0.000
place-labial	0.070	0.132	0.533	0.594
place-velar	-0.078	0.113	-0.694	0.488
V-/i/	0.095	0.091	1.036	0.300
V-/u/	0.111	0.091	1.213	0.225
closure duration	-2.563	0.555	-4.614	0.000
block 2	-0.111	0.082	-1.359	0.174
block 3	-0.167	0.082	-2.042	0.041
block 4	-0.177	0.082	-2.153	0.031
place-labial:V-/i/	-0.146	0.185	-0.790	0.429
place-velar:V-/i/	0.109	0.159	0.684	0.494
place-labial:V-/u/	-0.060	0.181	-0.330	0.741
place-velar:V-/u/	-0.067	0.156	-0.426	0.670
Num. obs.	1455			
Num. groups: word	69			
Var. word (intercept)	1.83e-09			
Num. groups: speaker	6			
Var. speaker (intercept)	2.93e-03			

Table A.4. English full model with word and speaker as random intercepts and logit link for beta regression. Call: voicing percent ∼ voicing + place × V + closure duration + block + (1 | word) + (1 | speaker)

	Estimate	Std. Error	z	p
(Intercept)	1.584	0.374	4.232	0.000
voicing-voiceless	-1.550	0.084	-18.389	0.000
place-labial	-0.117	0.156	-0.752	0.452
place-velar	-0.465	0.154	-3.028	0.002
V-/i/	0.432	0.154	2.807	0.005
V-/u/	0.371	0.167	2.219	0.026
closure duration	-9.957	1.212	-8.213	0.000
block 2	0.067	0.086	0.776	0.438
block 3	-0.039	0.085	-0.461	0.645
block 4	0.055	0.085	0.651	0.515
place-labial:V-/i/	-0.049	0.217	-0.223	0.823
place-velar:V-/i/	0.265	0.223	1.189	0.234
place-labial:V-/u/	0.070	0.237	0.294	0.768
place-velar:V-/u/	0.259	0.228	1.137	0.256
Num. obs.	1561			
Num. groups: word	61			
Var. word (intercept)	0.032			
Num. groups: speaker	6			
Var. speaker (intercept)	0.633			

APPENDIX B

POLISH AND FRENCH SIBILANTS

B.1 Graphs for all speakers

Figure B.1. COG contrasts in French: Speaker 03

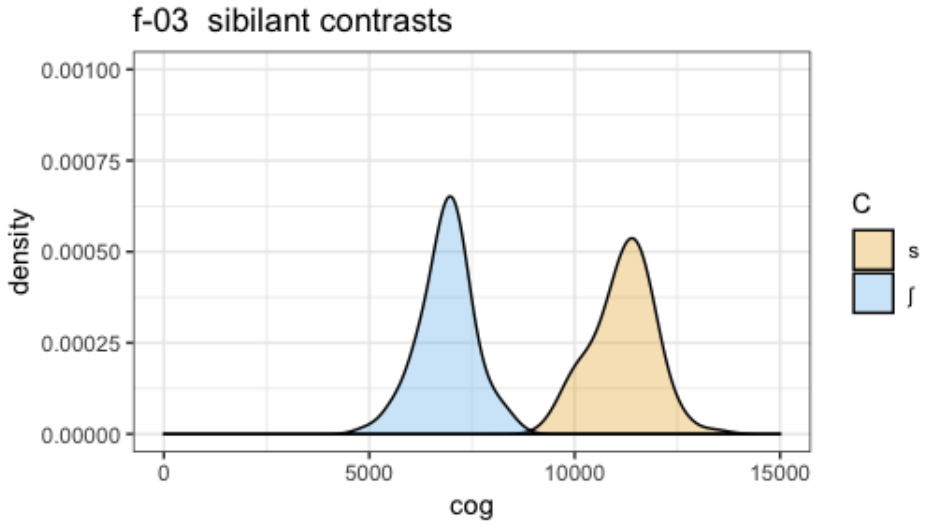


Figure B.2. COG contrasts in French: Speaker 04

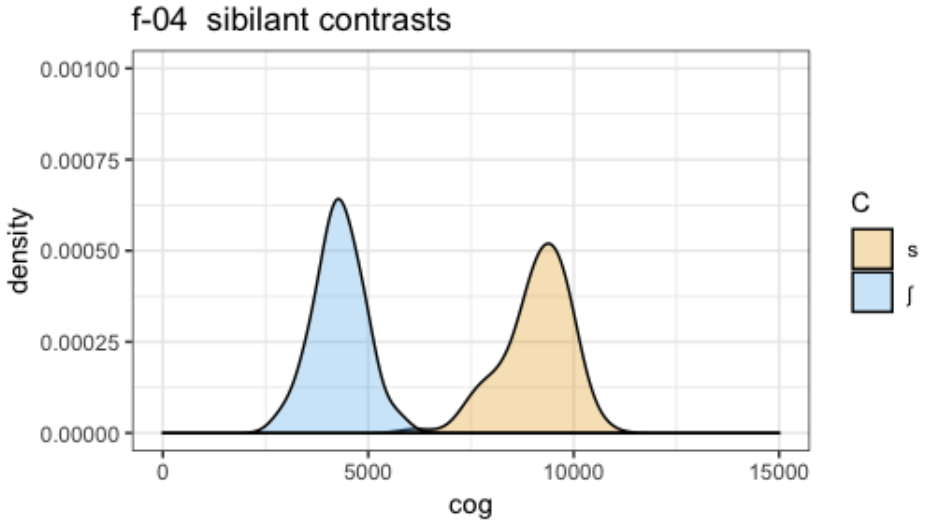


Figure B.3. COG contrasts in French: Speaker 05

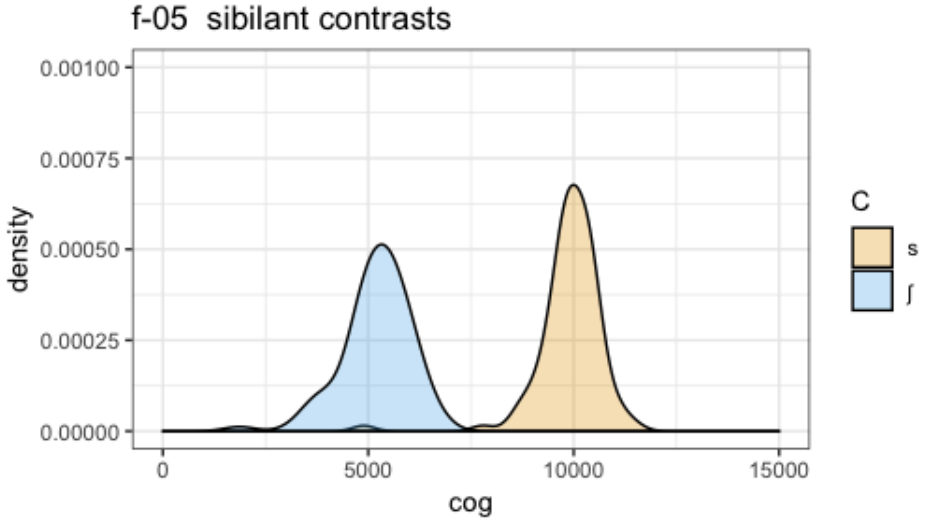


Figure B.4. COG contrasts in French: Speaker 06

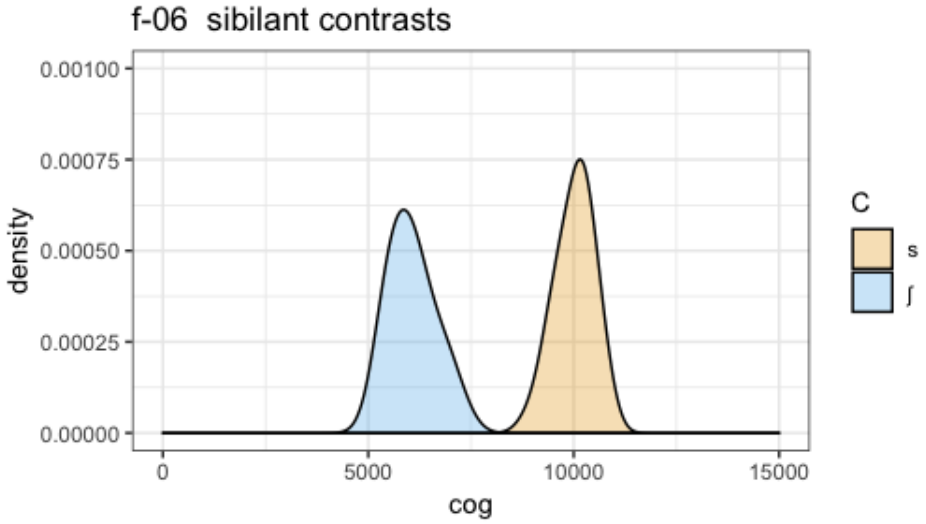


Figure B.5. COG contrasts in Polish: Speaker 03

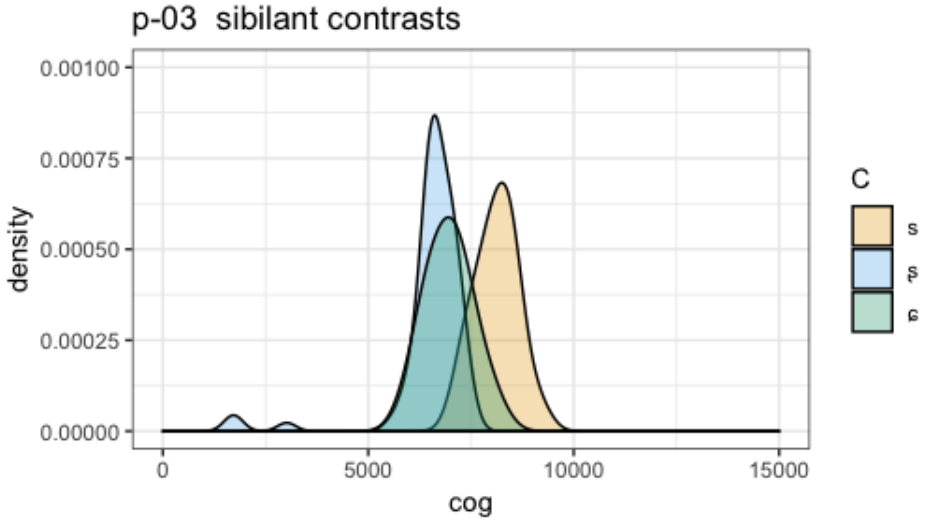


Figure B.6. COG contrasts in Polish: Speaker 05

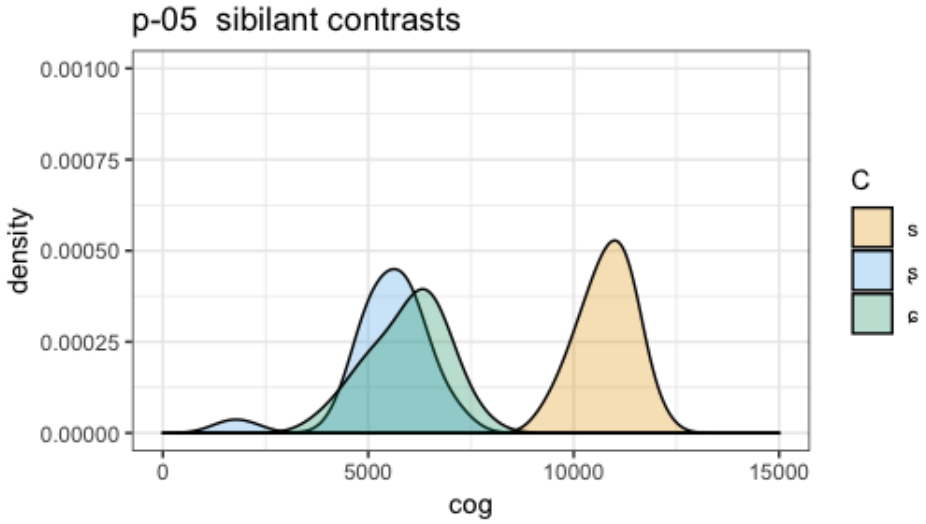


Figure B.7. COG contrasts in Polish: Speaker 06

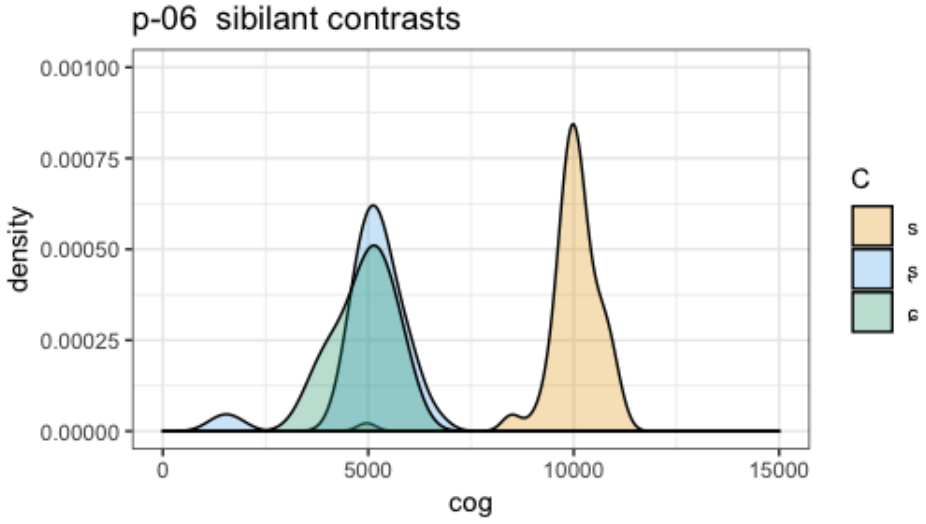


Figure B.8. F2 trajectories in French: Speaker 03

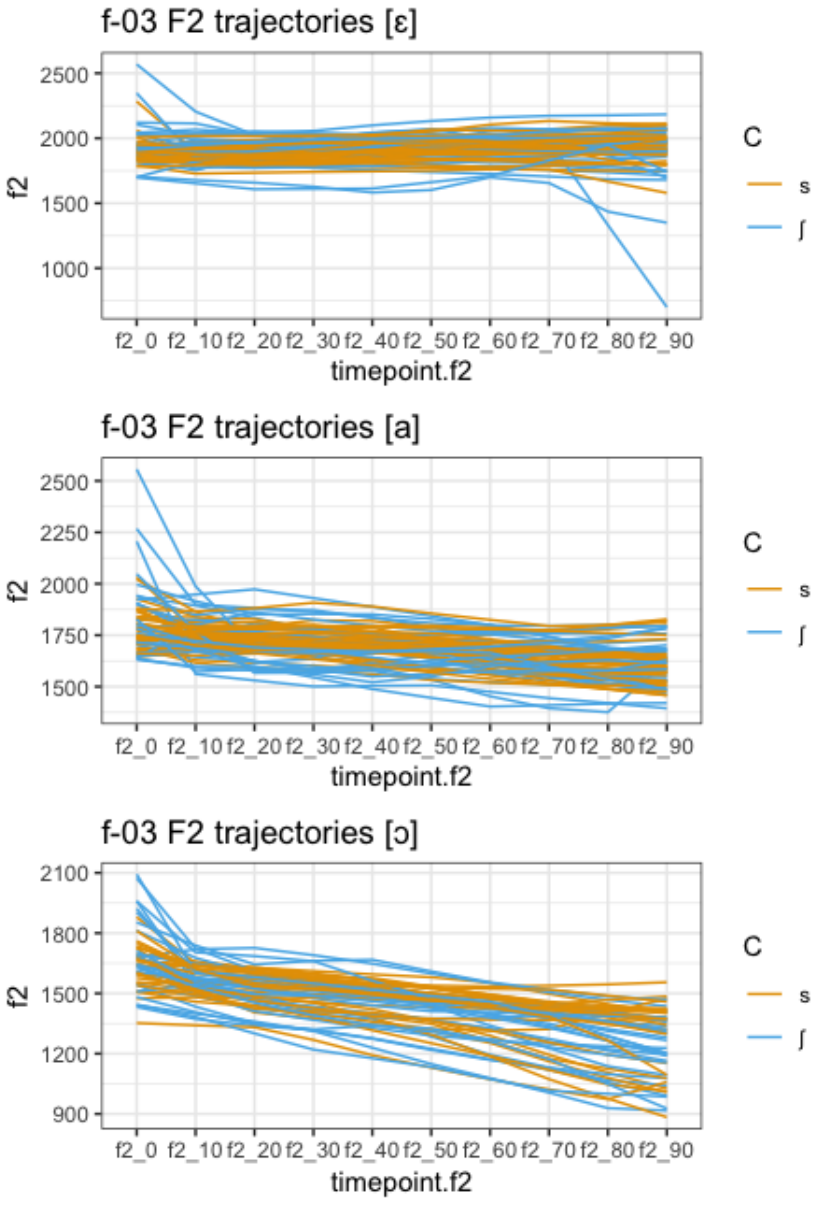


Figure B.9. F2 trajectories in French: Speaker 04

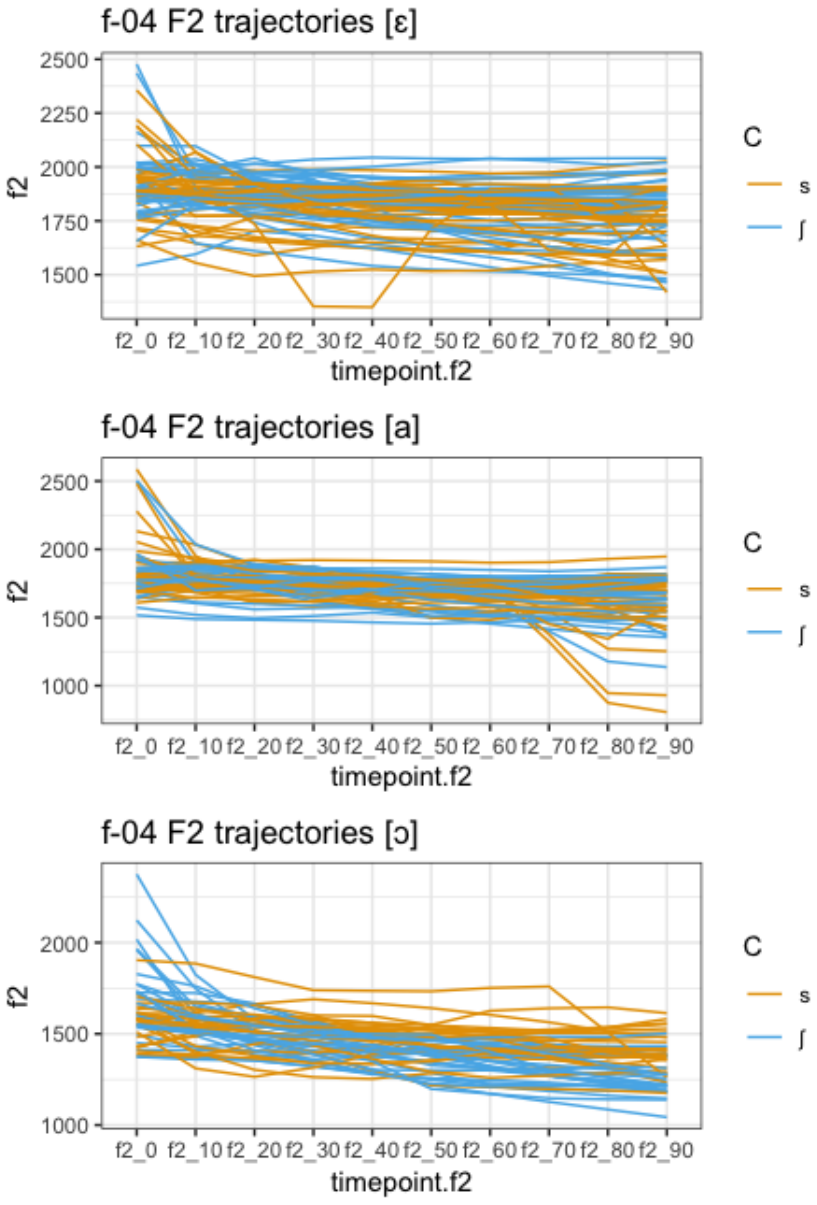


Figure B.10. F2 trajectories in French: Speaker 05

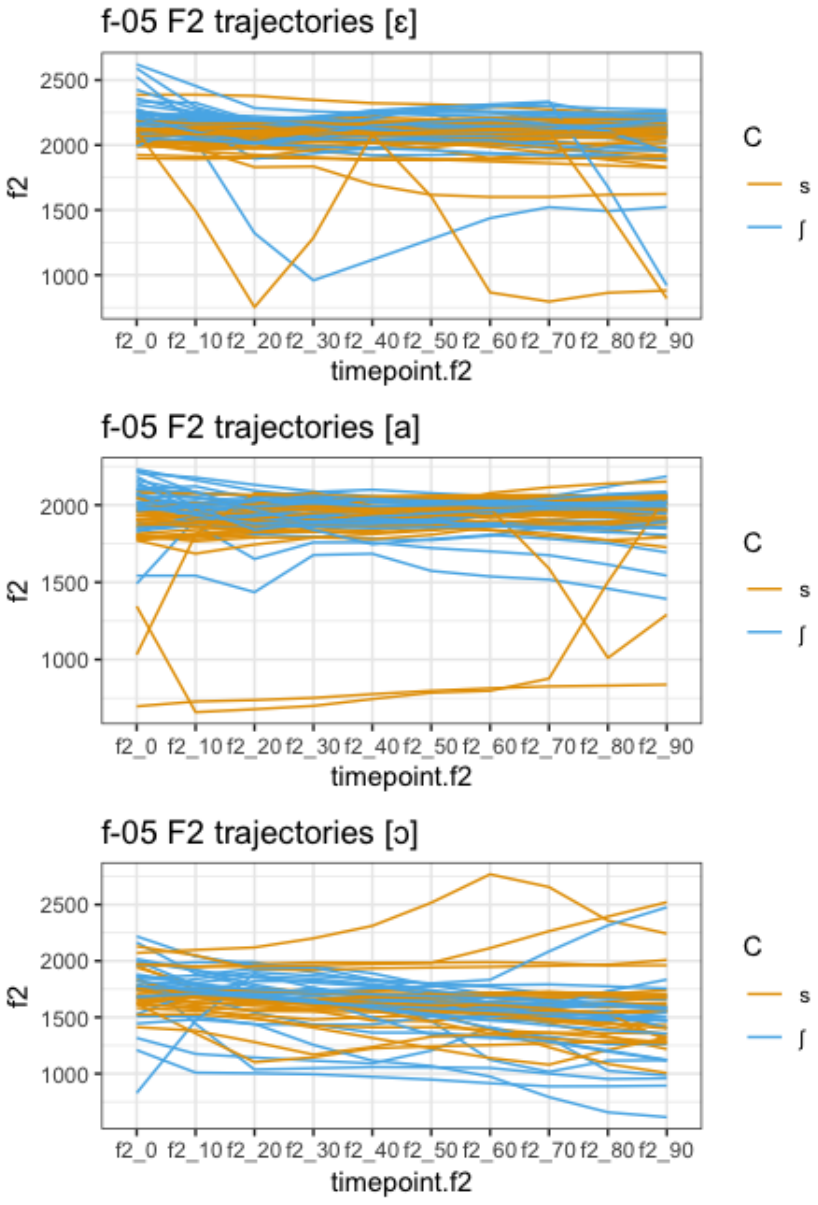


Figure B.11. F2 trajectories in French: Speaker 06

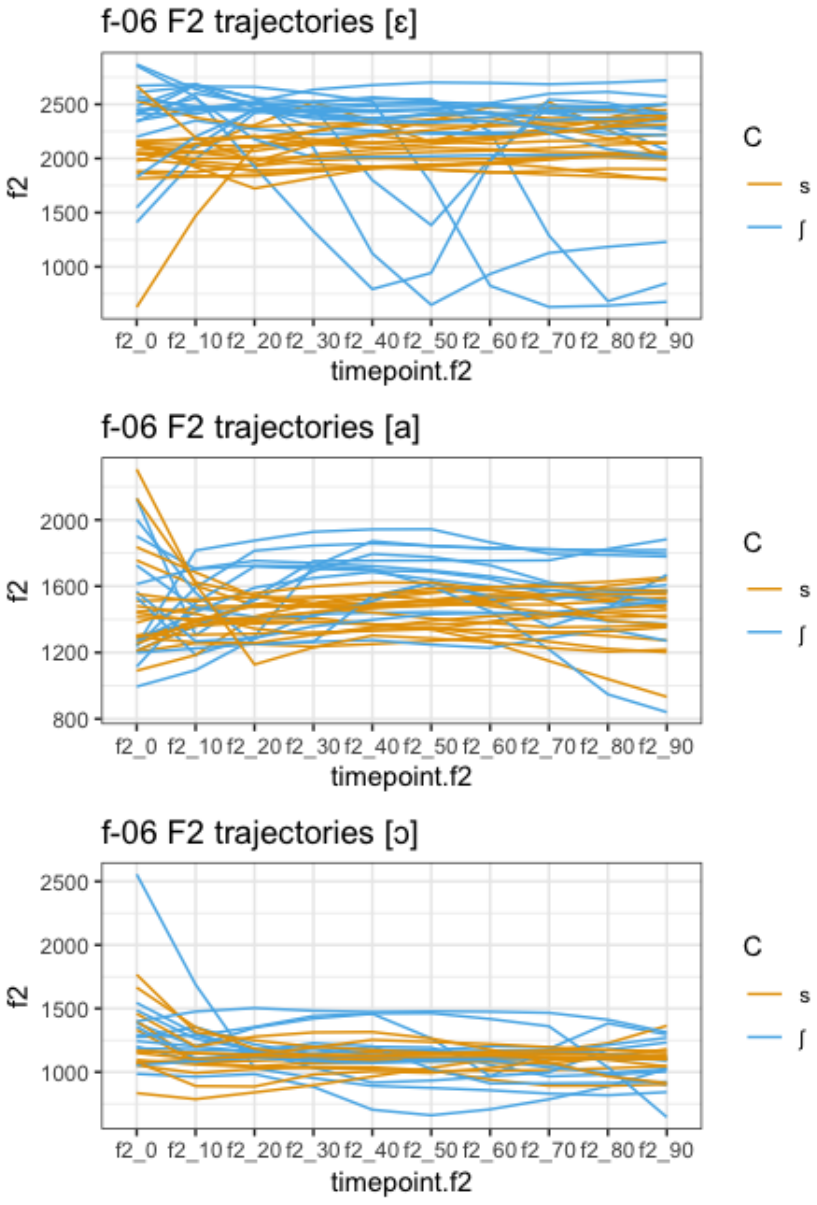


Figure B.12. F2 trajectories in Polish: Speaker 03

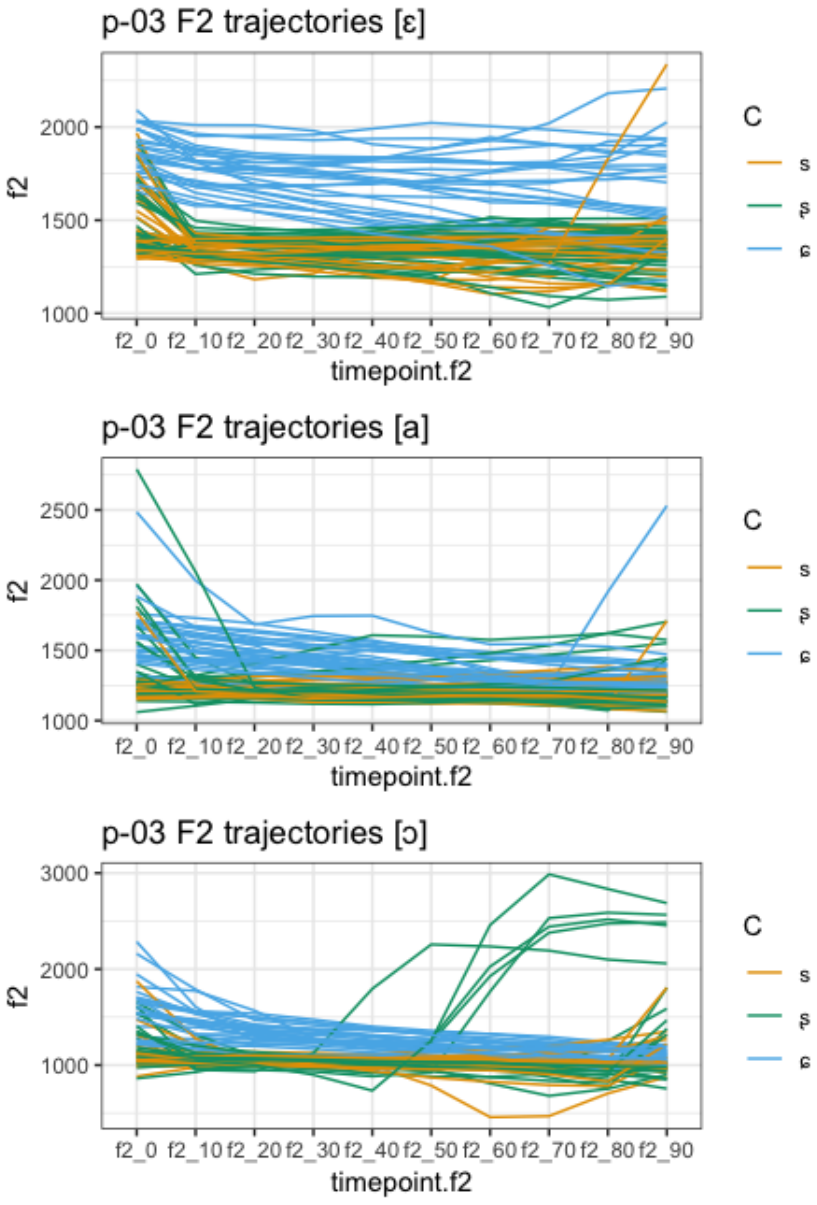


Figure B.13. F2 trajectories in Polish: Speaker 05

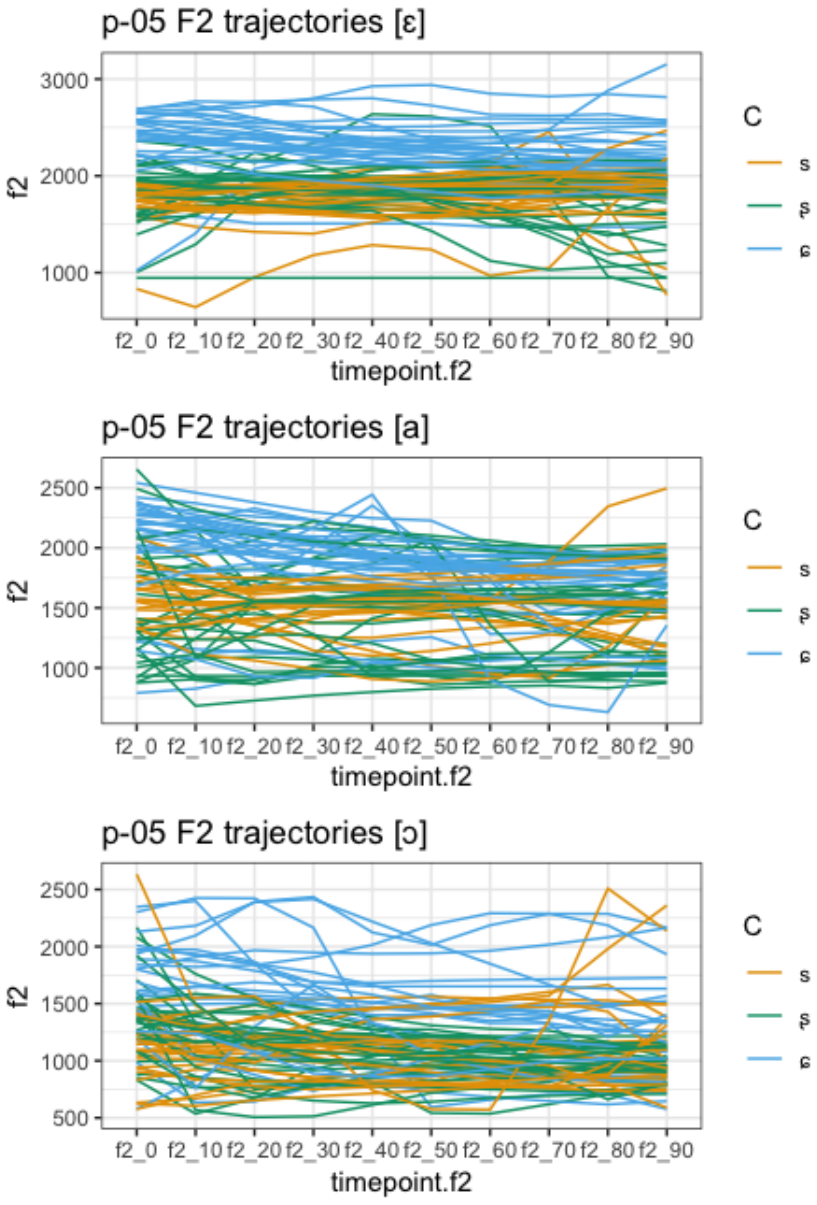
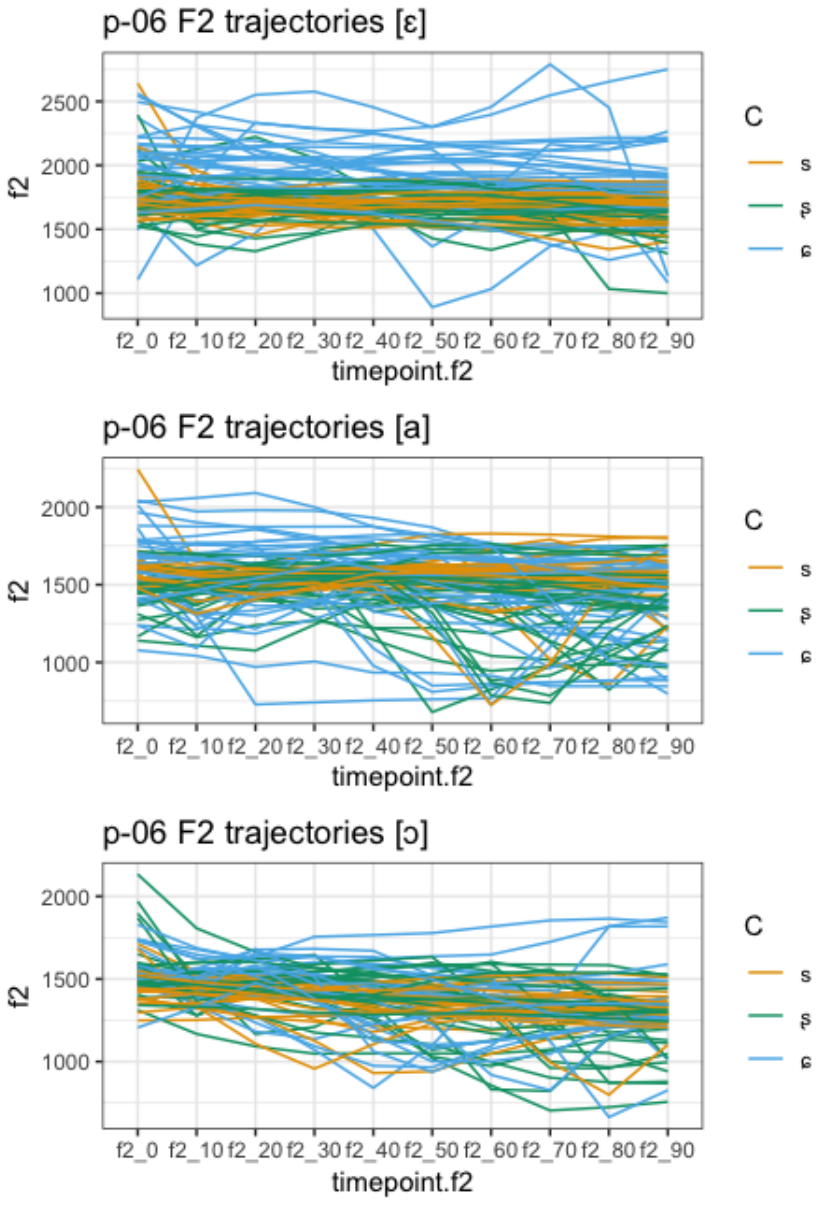


Figure B.14. F2 trajectories in Polish: Speaker 06



APPENDIX C

MANDARIN SIBILANTS

C.1 Graphs for all speakers

This appendix provides the two dimensional graphs showing sibilant contrasts for all speakers in the Mandarin experiment.

Figure C.1. Speaker m-02

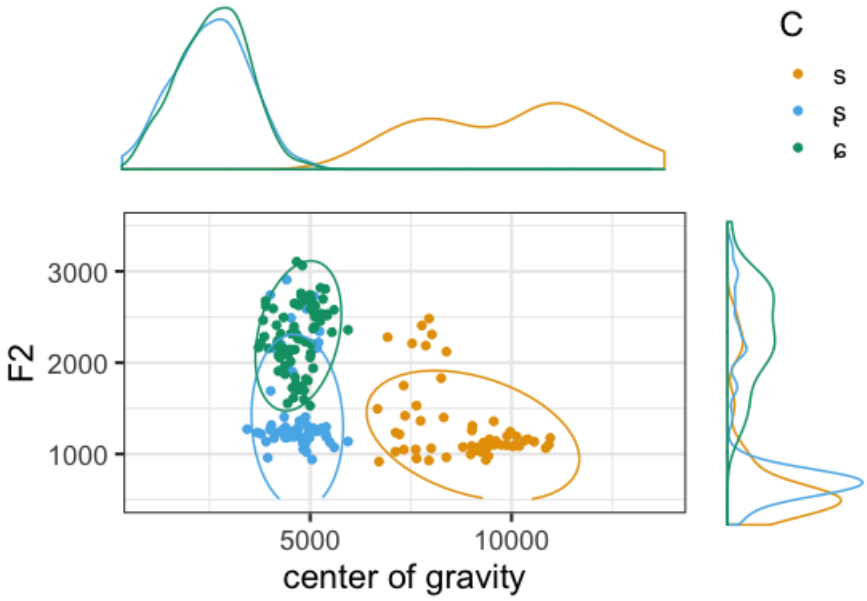


Figure C.2. Speaker m-03

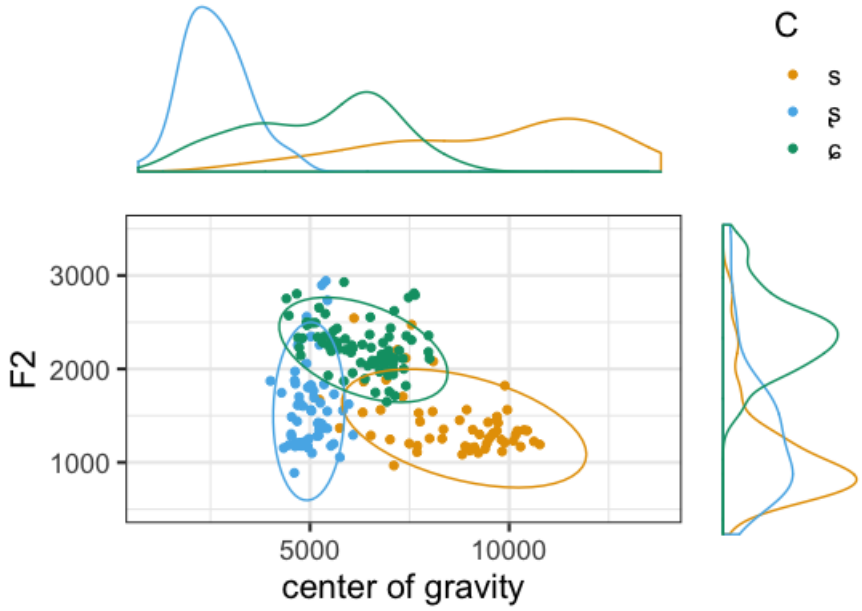


Figure C.3. Speaker m-06

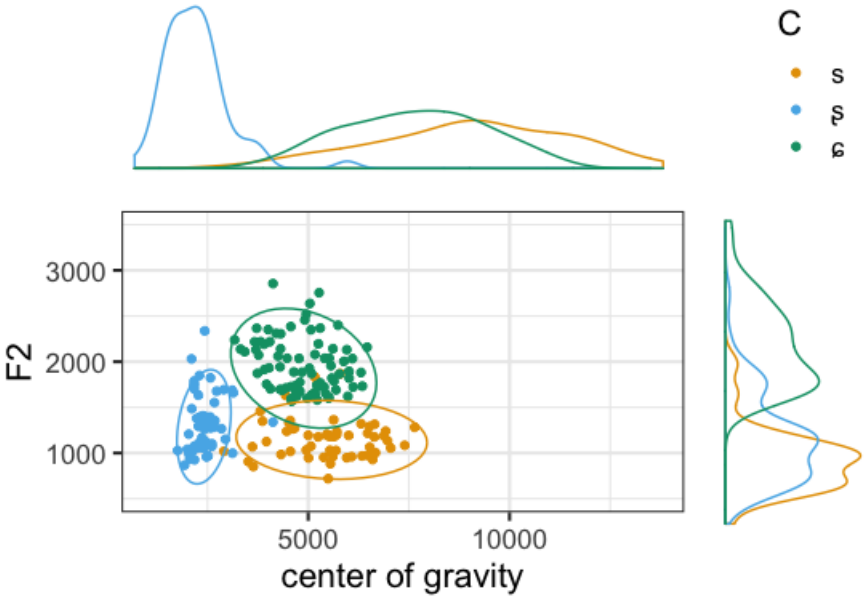


Figure C.4. Speaker m-07

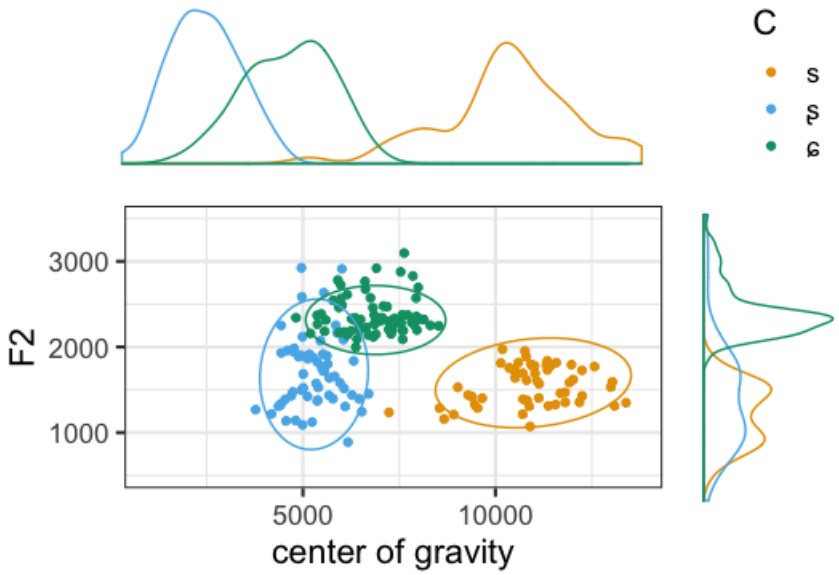


Figure C.5. Speaker m-08

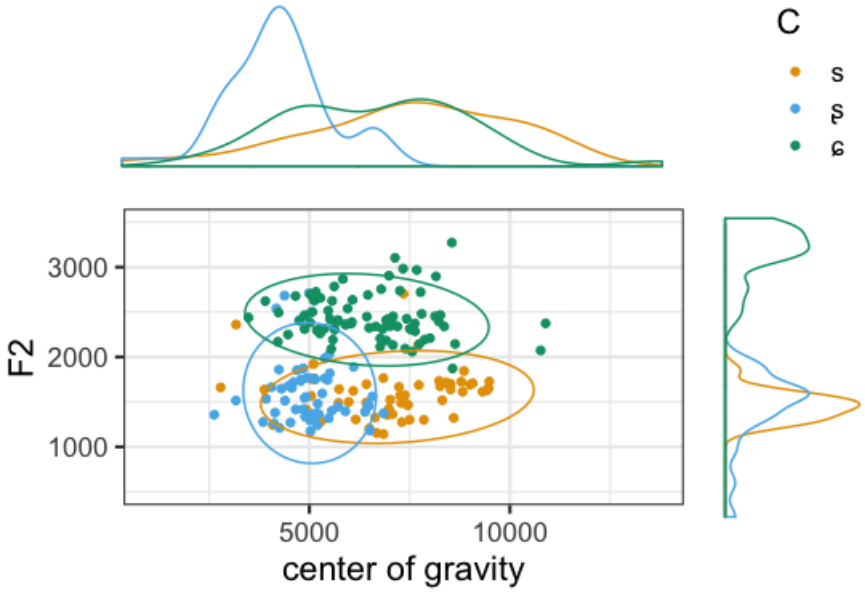


Figure C.6. Speaker m-09

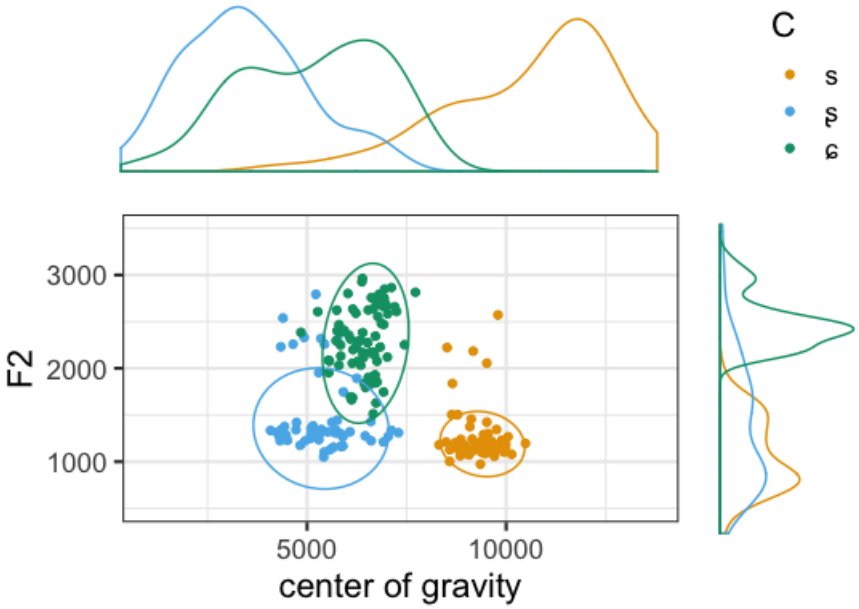


Figure C.7. Speaker m-15

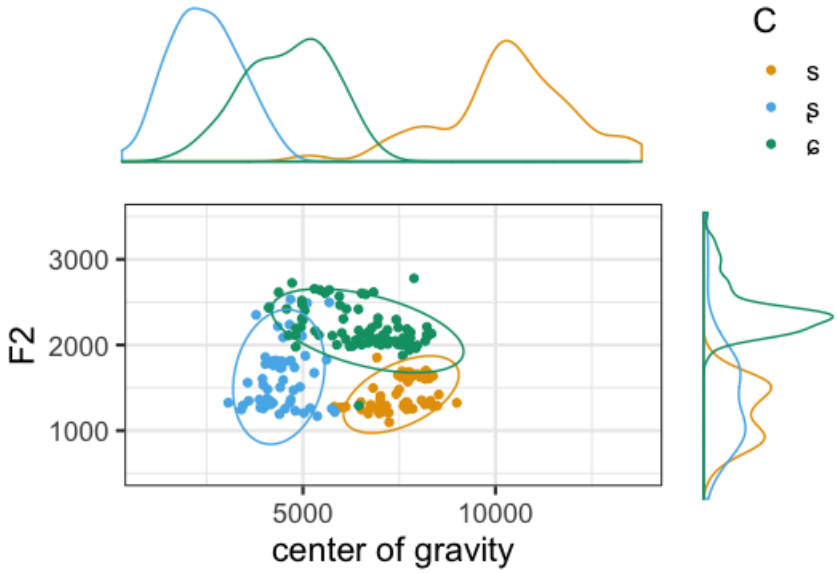


Figure C.8. Speaker m-17

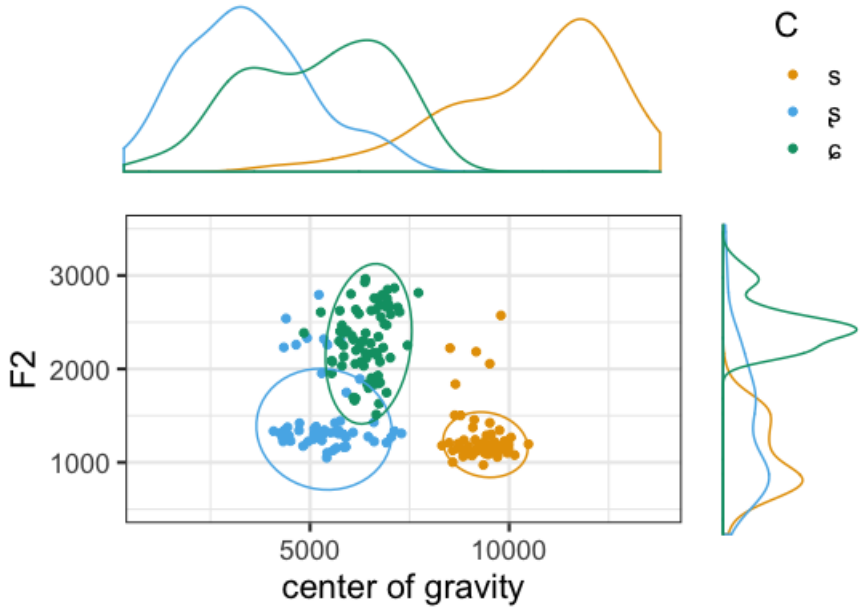
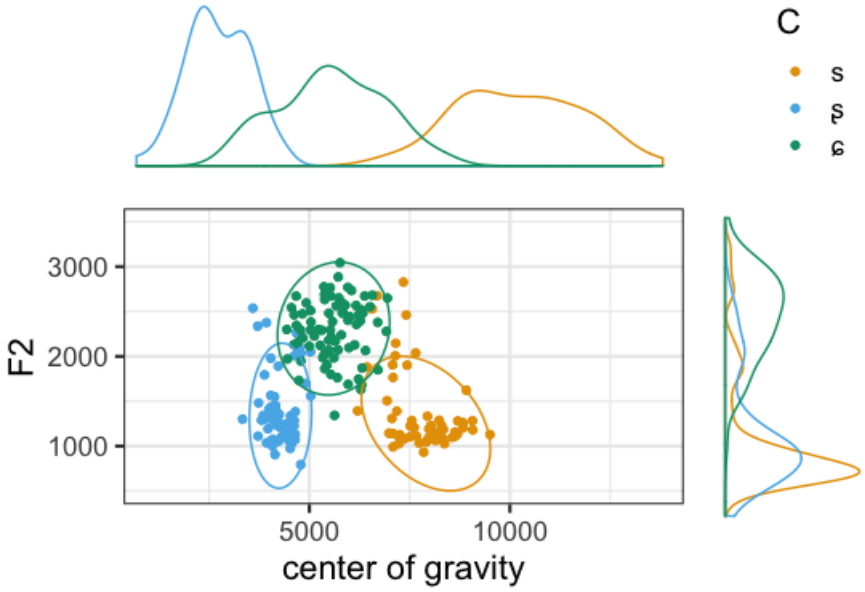


Figure C.9. Speaker m-19



C.2 Alternative analyses for cue weight

In Chapter 4, the Mandarin results showed that variation in F2 (onset of second formant of the following vowel) increases with degree of contrast along the COG (spectral center of gravity) dimension across speakers. In that analysis, I used LDA to quantify contrast on the COG dimension, taking the LDA coefficients as cue weights. In this appendix I re-consider those results, discussing use of the coefficients and comparing those results with alternatives obtained from the use of LDA error rate and JM distance.

C.2.1 Error vs. coefficients in LDA

There are multiple ways to compare the phonetic spaces of phonological inventories using LDA. In Chapter 4, I used the coefficients of linear discriminants. However, we might also consider using model error rate (as in the stop inventory case study) in place of the coefficients to measure contrast on the COG dimension. In this section, I show the same results using LDA error and discuss why the coefficients are preferable for the hypothesis being tested. To obtain error rates, LDA was performed with COG as the sole predictor variable and error rates were calculated by dividing the number of misclassified tokens by the total sample size. As in the stop inventory example, the training and testing data sets were identical.

Figure C.10 shows the same results displayed in Figure 4.12, but using the LDA error rate instead of the coefficients as the measure of COG separability. I show Figure 4.12 again here for reference (as Figure C.11). Higher coefficients indicate more contribution of the COG dimension to separability whereas lower error rate indicates more separability when COG is the sole predictor. If error was providing exactly the same information about contrast as the coefficients, we would expect to see the reverse trends we see in Figure C.11. In Figure C.10 (error results), there seem to be potential negative trends for the alveolar and alveopalatal sibilants and

Table C.1. Fixed effects table for linear mixed effects regression. Dependent variable: within-category within-vowel F2 variation, Predictors: COG separability (LDA error), C, V, C×COGsep interaction, random intercepts for speaker. Intercept is [ʂa].

Fixed effects	Estimate (se)	t	p
(Intercept)	10.29(2.63)	3.90	$< 0.001^{**}$
COGsep (LDA error)	6.60(12.95)	0.51	0.612
C-/s/	0.50(3.54)	0.14	0.889
C-/ɕ/	−1.93(3.54)	-0.54	0.590
V-/u/	12.30(1.92)	6.40	$< 0.001^{***}$
COGsep × C-/s/	−19.53(17)	-1.15	0.257
COGsep × C-/ɕ/	−34.91(17)	-2.05	0.046^{*}

no visible trends for the retroflex in either vowel context. In the /a/ context there is one particular speaker, m-02, who has high COG error and does not seem to pattern with the potential negative trend. This speaker does seem to pattern with the group when the LDA coefficients are used.

In Table C.1, I provide the results from a mixed effects linear regression model relating LDA error to F2 variation (analogous to the results in Table 4.4 but using LDA error rate instead of LDA coefficients as a predictor). This model still shows the effect of more variation in the /u/ vowel context, but the effect of COG separability is different. There is still no significant effect for the effect of COG separability on F2 variation for the intercept /ʂ/. In the model using the coefficients for cue weight, we observed a significant effect of the COG coefficients for both /s/ and /ɕ/ (as determined by the interactions; see Chapter 4 for a more detailed explanation of this interpretation). In this model, we do not see a significant effect of LDA error for the alveolar sibilant, but we do see a significant effect for the alveopalatal sibilant (though it is marginal at $p = 0.046$).

The differences in model outcomes between the model in Table 4.4 and the model in Table C.1 are due to the relative differences in LDA error vs. LDA coefficients. LDA coefficients provide a measure of the relative weight of the COG dimension

Figure C.10. COG separability (LDA error rate) and F2 variation across speakers: Alveolar sibilant in top panel, alveopalatal sibilant in middle panel, retroflex sibilant in bottom panel.

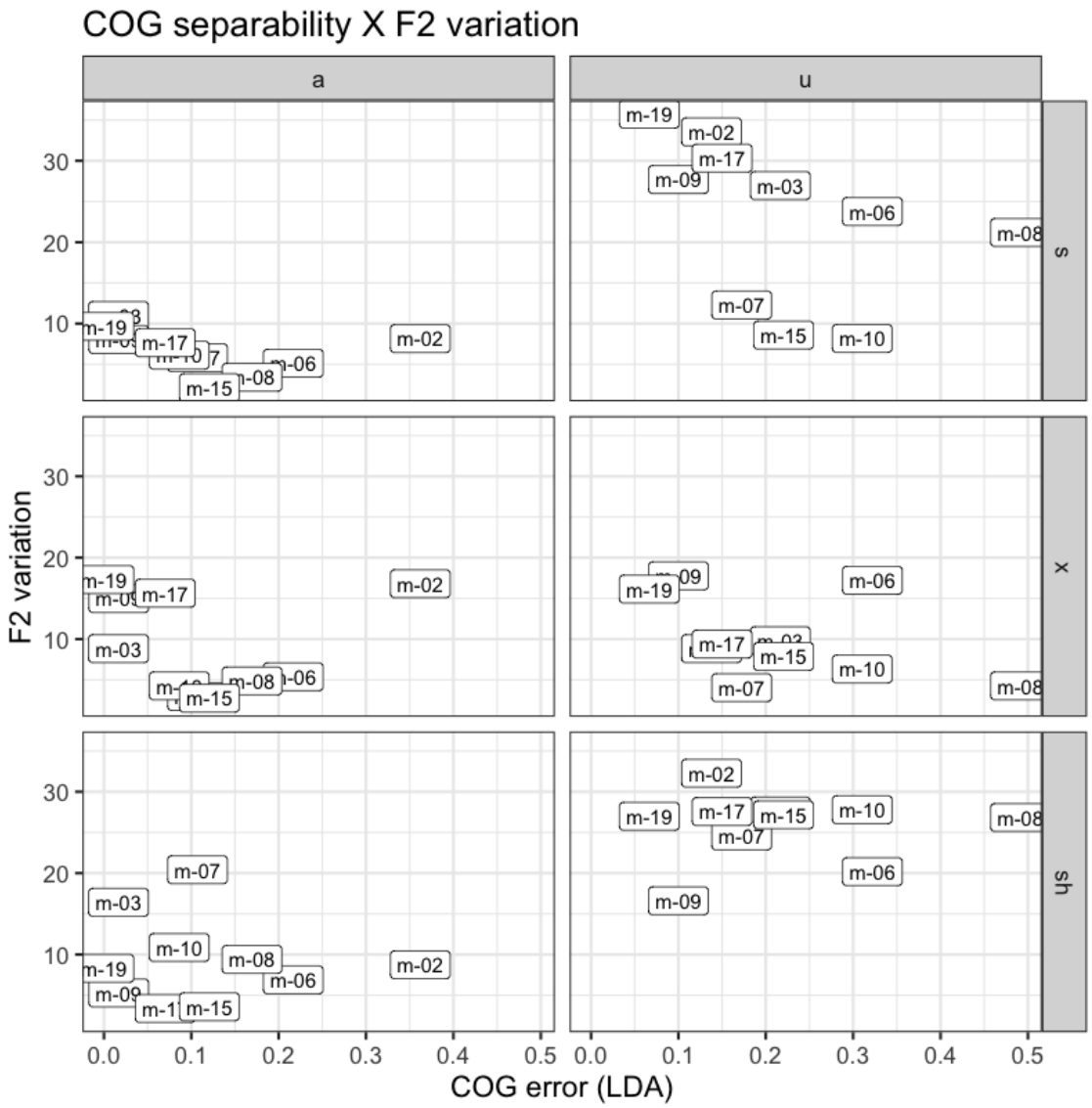


Figure C.11. COG separability (LDA coefficients) and F2 variation across speakers: Alveolar sibilant in top panel, alveopalatal sibilant in middle panel, retroflex sibilant in bottom panel (reprinted here from Chapter 4 for comparison with Figure C.10).

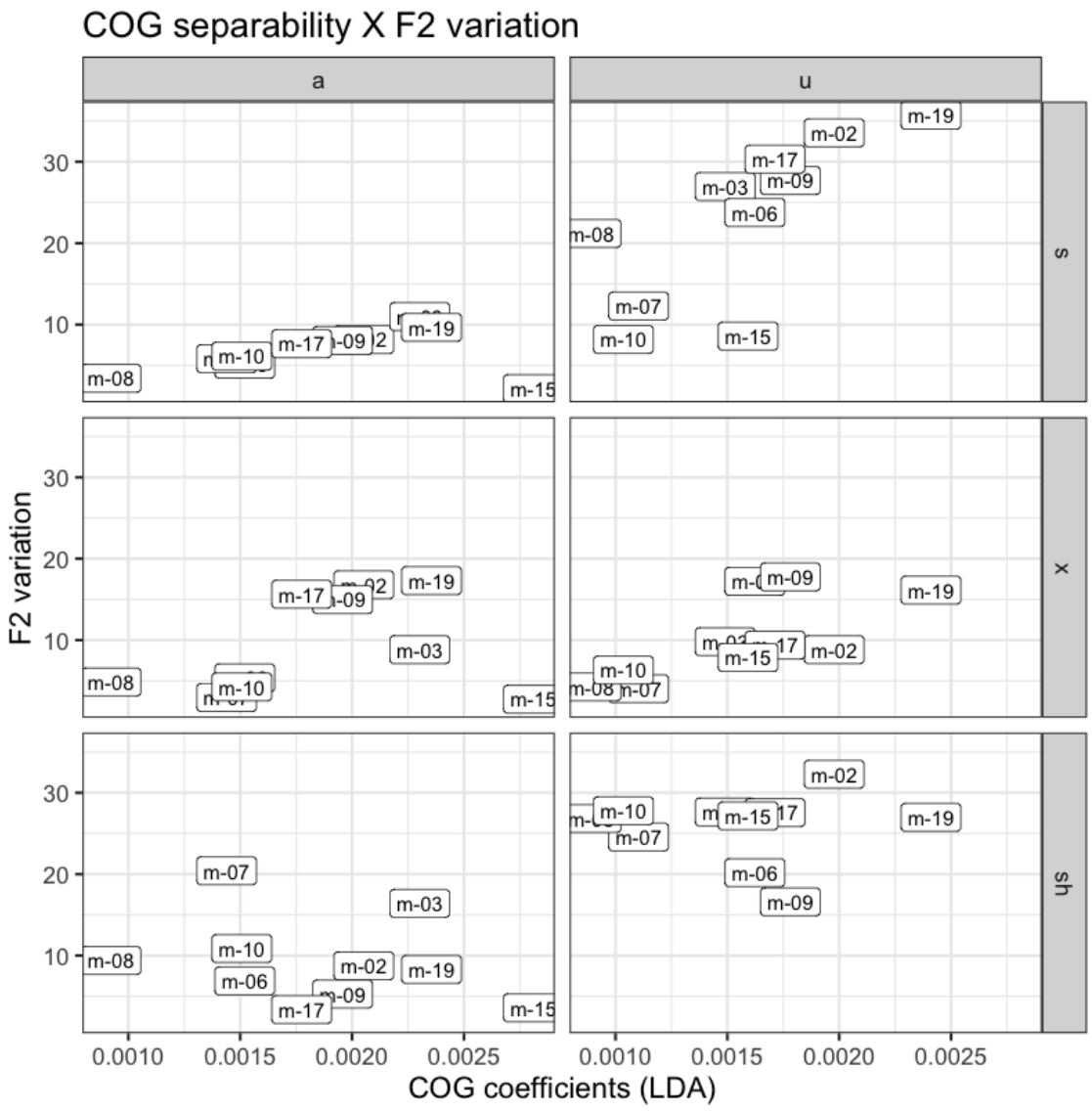
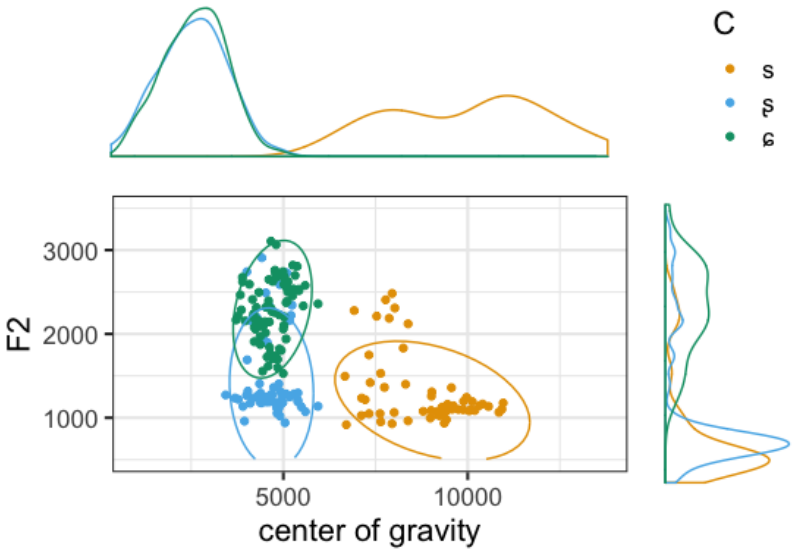


Figure C.12. Example speaker: 2 distinct categories on COG dimension



in determining separability, whereas LDA error provides a measure of the overall classification accuracy of the model with COG as the sole predictor. These values are intuitively related, but they are not perfectly correlated.

I illustrate the difference between LDA error and coefficients by considering the speaker with the biggest difference in results between the LDA coefficients and LDA error analyses. In Figure C.10, there is one speaker, m-02, who consistently does not pattern with the rest of the group in the /a/ context (left panel). This speaker has the highest error rate in that context. The pattern with the rest of the speakers appears to be a negative trend between error and F2 variation for /s/ and /ɕ/. In Figure C.12 we see the raw data for this speaker. The COG density plot above the x-axis shows good separability between the /s/ category and the other two sibilants. However, the distributions for /ɕ/ and /ʂ/ are almost entirely overlapping. This overlap results in a high overall error rate in classification when using COG as the only predictor. The high error rate does not reflect the fact that there is almost perfect separation between /s/ and the other two sibilants on the COG dimension. This separation results in this speaker having a relatively higher COG coefficient value.

LDA error provides a more direct measure of linear separability, effectively corresponding to amount of category overlap on the COG dimension. This differs from the LDA coefficients, which provide a measure of how much the COG dimension contributes to overall separability. Error rate provides a better metric for comparing overall category separability, while the coefficients provide a better metric for comparing strength of COG as a predictor. In the stop inventory example, we were mainly interested in how separable the categories were overall and were not interested in the relative contribution of each of the predictor dimensions (F1 & F2). In that case, the overall error rate of the LDA model was more appropriate for comparing the inventories.

However, the main question of interest with the Mandarin sibilants does not deal with overall separability across multiple dimensions, but rather the degree of contrast on a particular dimension (COG). As the sibilants contrasts are multidimensional, assessing the contrast on one of those dimensions is better approximated by the strength of that dimension as a predictor rather than the overall error rate of an LDA model. Therefore, LDA coefficients are preferable to error for this particular analysis. This is in line with previous work in phonetics which uses LDA coefficients as a measure of relative cue weight in production (see Chapter 4).

C.2.2 LDA coefficients vs. JM distance

In Chapter 4, the main result was that F2 variation increases with COG contrast (as defined by the coefficients of linear discriminants from LDA) for /s/ and /ɕ/. This relationship is not significant for any of the sibilants when COG dispersion from JM distance is instead used as a metric of contrast. One reason for this is that the dispersion metric continues to increase with mean distance between categories, even when the categories are perfectly separable. I will showcase this difference with results

from example speakers and revisit the results from Chapter 4 using dispersion instead of LDA coefficients as a predictor of F2 variation.

COG dispersion was calculated by applying the JM distance to calculate the acoustic distances between each of the three sibilant categories for each speaker in each vowel context.[1] Because these calculations were done over a single dimension (the COG dimension), triangle area was not appropriate for calculating a dispersion measure. Instead, the distances between each pair of sibilants were summed to create an aggregate measure of category dispersion for each speaker in each vowel context.

The difference between the dispersion metric and the separability metric is well illustrated by examining speakers whose results change between the two methods. Speaker m-07 ranks lower among all speakers in separability from LDA, but higher in dispersion from JM distance. Raw data from this speaker is given in Figure C.13 (top panel). Compare this with the data from another speaker, m-06 in the bottom panel, who ranks higher than m-07 in separability but lower than m-07 in dispersion in both vowel contexts.

Across all speakers, the distance between /ʂ/ and /s/ for m-07 is one of the highest mean-to-mean distances between any two categories on the COG dimension, which contributes to the high dispersion score. The JM distance does incorporate within-category variation in addition to mean-to-mean distance, but the ratio of between-category to within-category variation between /s/ and /ʂ/ for m-07 is still much higher relative to other speakers, resulting in the higher dispersion score. Despite this, there is still considerable overlap between /ʂ/ and /ɕ/ on the COG dimension for this speaker, which leads to the relatively lower COG coefficients. Consider speaker m-06 (bottom panel of Figure C.13), who ranks lower than m-07 in dispersion. This speaker

[1]Distances were calculated between /s/-/ʂ/, /s/-/ɕ/, and /ʂ/-/ɕ/ in /a/ and /u/ contexts.

Figure C.13. Sibilant categories in COGxF2 space for two speakers. m-07 (top panel) ranks lower than m-02 (bottom panel) in separability but higher in dispersion.

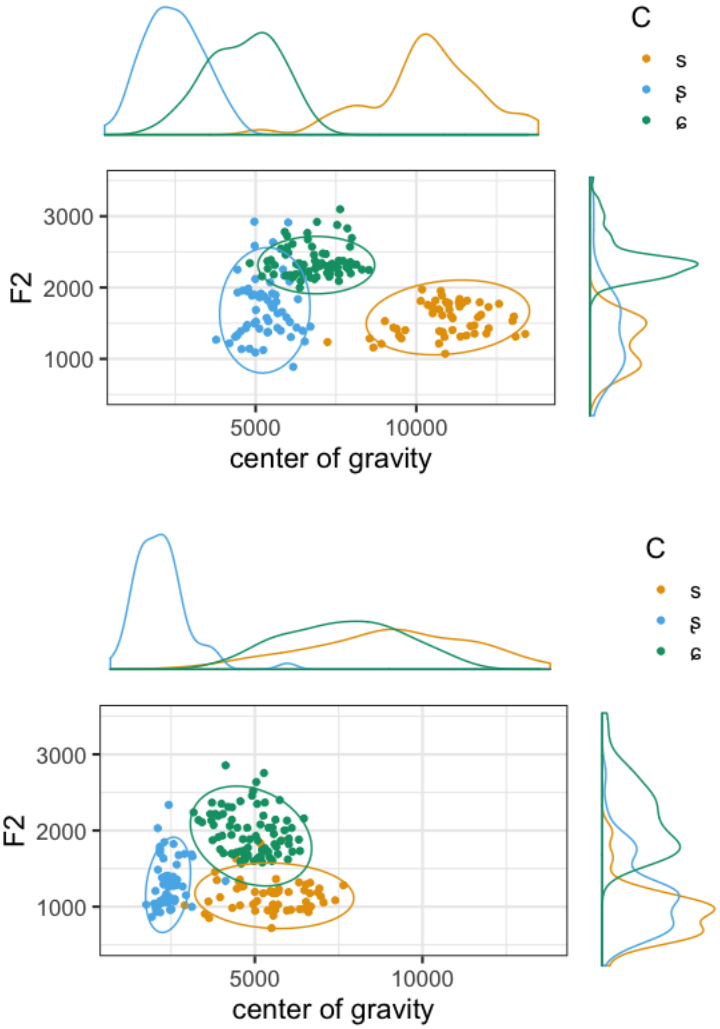


Table C.2. Fixed effects table for linear mixed effects regression. Dependent variable: within-category within-vowel F2 variation, Predictors: COG dispersion (JM distance), C, V, C×COGdisp interaction, random intercepts for speaker. Intercept is [ʂa].

Fixed effects	Estimate (se)	t	p
(Intercept)	10.30(7.51)	1.37	0.177
COGdisp (JM distance)	0.24(2.16)	0.11	0.911
C-/s/	−0.66(9.58)	-0.69	0.946
C-/ɕ/	−22.86(9.58)	-2.39	0.021^{*}
V-/u/	12.04(1.81)	6.64	$< 0.001^{***}$
COGdisp × C-/s/	−0.32(2.85)	-0.11	0.911
COGdisp × C-/ɕ/	4.68(2.85)	1.64	0.110

exhibited smaller between-category distance but almost perfect separability between /ʂ/ and the other two sibilants, leading to higher COG coefficients.

The results comparing COG dispersion and F2 variation are given in Figure C.14. This is analogous to Figure 4.12 (but with JM distance as the metric of COG separability instead of the LDA coefficients), which is printed again here for reference as Figure C.15. The original hypothesis predicts a positive correlation between degree of COG contrast and F2 variation across speakers. When COG contrast is quantified as category dispersion with JM distance, we do not see any significant trend. The data for the alveolar sibilant in the /u/ context are potentially approaching a positive trend across speakers, but this is not significant when modeled with linear mixed effects regression (Table C.2). None of the expected interactions are significant, indicating no significant relationships between COG dispersion and F2 variation across speakers.

C.2.2.1 Interim discussion: Separability and dispersion in Mandarin sibilants

In the case of the Mandarin sibilants, we are more interested in how the COG dimension is used contrastively by each speaker. The dispersion metric does not

Figure C.14. COG dispersion and F2 variation across speakers: Alveolar sibilant in top panel, alveopalatal sibilant in middle panel, retroflex sibilant in bottom panel.

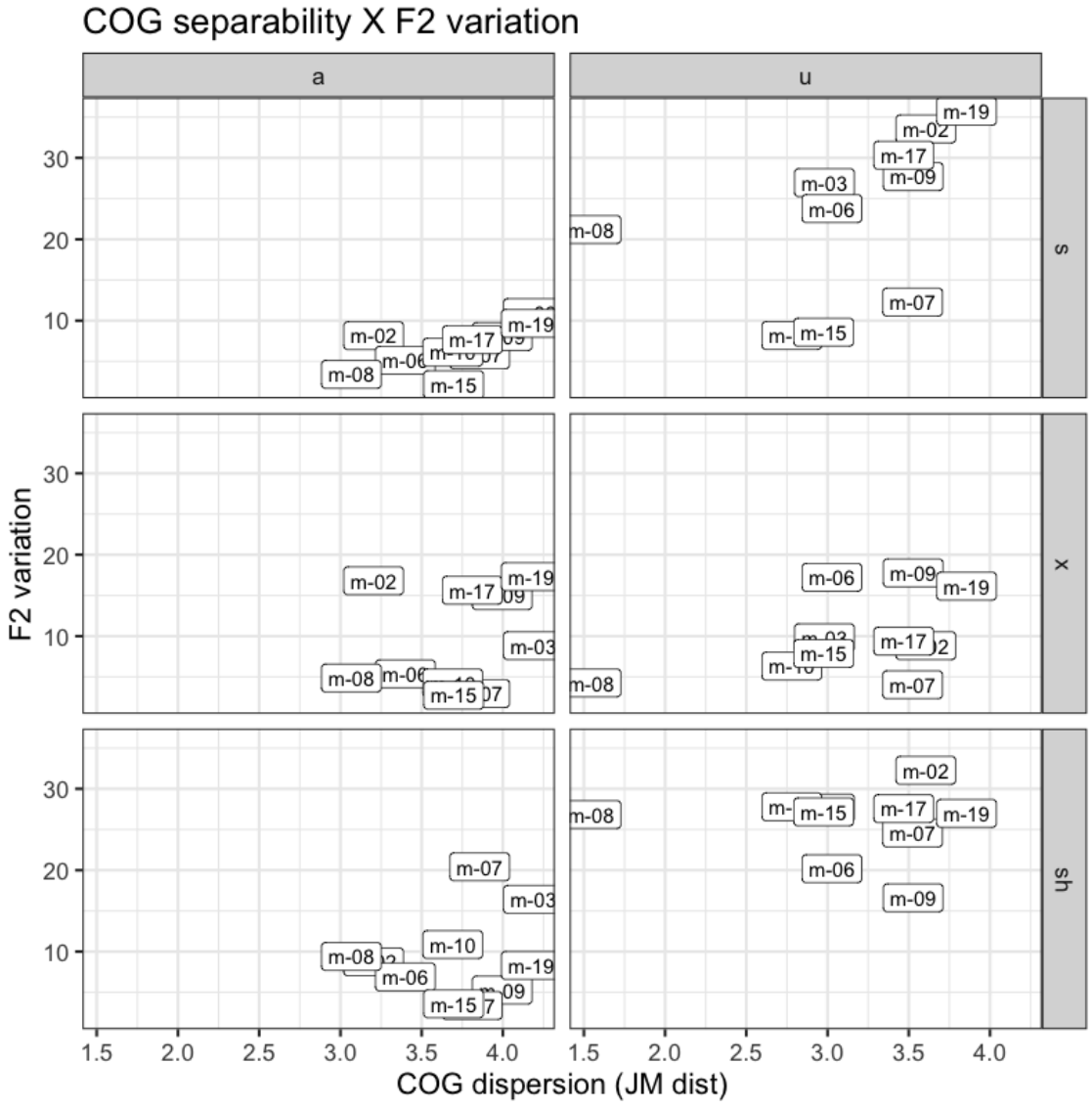
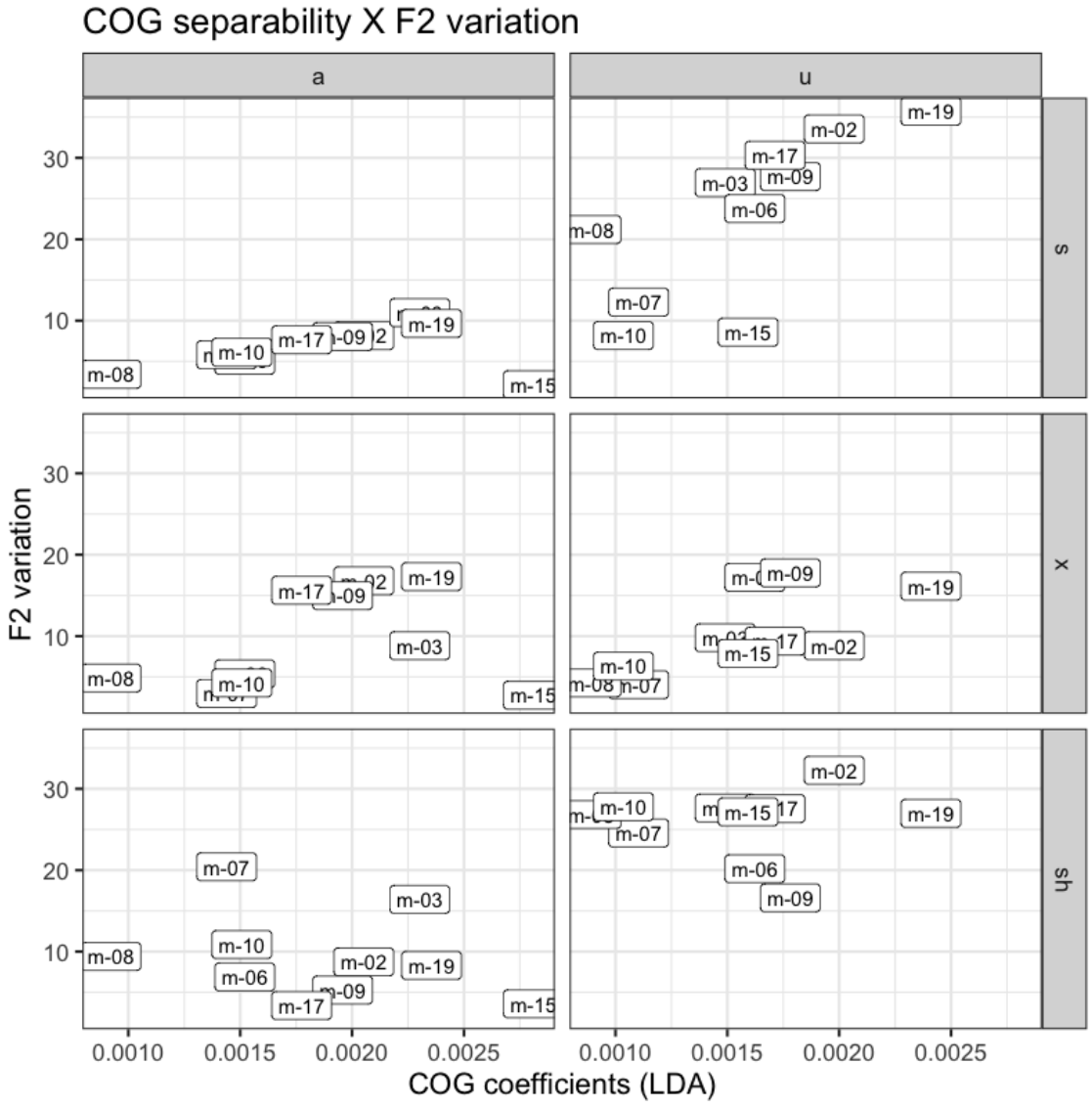


Figure C.15. COG separability and F2 variation across speakers: Alveolar sibilant in top panel, alveopalatal sibilant in middle panel, retroflex sibilant in bottom panel.



distinguish between sufficient and maximal category dispersion, a notion introduced to Dispersion Theory in Lindblom (1986), which is also relevant here. Phonetic categories need only be *sufficiently* dispersed for phonological contrast. The JM distance provides higher values for categories which are maximally dispersed relative to categories which are sufficiently dispersed for contrast. The coefficients from LDA provide a metric of COG's contribution to overall contrast and are preferable for this analysis. The results shown here demonstrate that acoustic dispersion, category separability, and cue weight along a single dimension all provide different results despite the fact that they seem intuitively related.

In this appendix, I have reviewed LDA error, LDA coefficients, and JM distance as ways of quantifying relationships between contrastive phonological categories in acoustic space. LDA error is best used in cases where the relevant hypothesis deals with overall category separability (especially if across multiple phonetic dimensions) as in the stop inventory example, LDA coefficients are best used for cases where the relevant hypothesis deals with the relative contribution of multiple phonetic dimensions to overall category separability as in the Mandarin sibilants example, and the JM distance is best used for cases where the relevant hypothesis deals with overall category spread, and is particularly relevant for testing hypotheses related to Dispersion Theory.

APPENDIX D

METRICS FOR DISPERSION IN STOP INVENTORIES

D.1 LDA analysis

To determine relative differences in separability between stop inventories, I ran LDA models for every possible 3 stop inventory. Calculations were done in R (R Core Team, 2013) using the MASS package (Ripley et al., 2013). In each inventory, the set of observations was all stop tokens generated by the model at the places of articulation included in each inventory. The predictor variables were F2 and F3 values and the dependent variable was the POA category labels.

The error rate of LDA classification provides one way of quantifying how separable the categories are. This error rate reflects how well the linear combination from LDA separates the data into the labeled categories. Error rates for the stop inventory data were calculated by dividing the number of misclassified tokens by the total sample size. These calculations were done with the training set as the testing set, a method which typically leads to lower error rates relative to partitioning a subset of the data for testing (James et al., 2013). This is not an issue for the present analysis since we are only concerned with relative differences in error rate, not with the error rate values themselves.

In Table D.1, I provide LDA error results for selected inventories. The results differ from both sets of dispersion results. The inventory with the smallest error (which can be considered the most linearly separable inventory) is the /bilabial coronal uvular/ inventory. The inventory which is most dispersed according to both metrics examined here, /coronal velar epi-pharyngeal/, is only the ninth most separable. The typolog-

Table D.1. LDA classification error results: $< F2, F3 >$ space

	POA1	POA2	POA3	LDA error rate
1	bilabial	coronal	uvular	0.0019
2	bilabial	uvular	epi-pharyngeal	0.0047
3	bilabial	palatal	uvular	0.0051
⋮				
9	coronal	velar	epi-pharyngeal	0.0140
10	coronal	uvular	velar	0.0190
11	bilabial	uvular	velar	0.0660
12	**bilabial**	**coronal**	**velar**	**0.0661**

ically common /bilabial coronal velar/ inventory is even less separable at number 11.

Intuitively, dispersed inventories have categories which are also easily separable. However, as these results have shown, separability (as defined by LDA error rate) and dispersion (as defined by JM distance) are not quantitatively identical. In this particular set of data, the uvular category is present in all of the five most separable inventories, but only one of the five most dispersed inventories. The uvular distribution has smaller within-category variance relative to the other places. This results in good separability even for categories which have similar means.

Metrics of dispersion such as JM and mean-to-mean distance provide a way of quantifying distance between categories. I have argued in Chapter 5 that this is best done with a measure of distance that includes within-category variance. If dispersion is understood as a ratio of within-category to between-category variance, dispersed inventories frequently have good separability and little overlap. However, this is not always the case. These results show that dispersion and separability, while related, are not quantitatively identical.

BIBLIOGRAPHY

Abramson, A. S., Lisker, L., 1967. Laryngeal behavior, the speech signal and phonological simplicity. In: Proceedings of the Tenth International Congress of Linguistics, Bucharest, vol. 4.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Alexander, J. A., 2010. The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems. Ph.D. thesis, Northwestern University.

Allen, J. S., Miller, J. L., DeSteno, D., 2003. Individual talker differences in voice-onset-time. The Journal of the Acoustical Society of America 113, 544–552.

Audacity Team, 1999-2014. Audacity(R): Free Audio Editor and Recorder. http://audacity.sourceforge.net/.

Baayen, R., 2008. Analyzing linguistic data: A practical introduction to statistics using R., vol. 37:2. Cambridge University Press.

Baayen, R. H., Davidson, D. J., Bates, D. M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language 59, 390–412.

Baker, W., Trofimovich, P., 2006. Perceptual paths to accurate production of L2 vowels: The role of individual differences. IRAL–International Review of Applied Linguistics in Language Teaching 44, 231–250.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., Treiman, R., 2007. The English lexicon project. Behavior research methods 39, 445–459.

Bang, H.-Y., Clayards, M., 2016. Structured variation across sound contrasts, talkers, and speech styles. Poster presented at LabPhon15: Speech Dynamics and Phonological Representation. Ithaca, NY .

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., Yoon, T.-J., 2018. The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. Journal of Phonetics 66, 120–144.

Barr, D. J., 2013. Random effects structure for testing interactions in linear mixed-effects models. Frontiers in Psychology 4, 328.

Bates, D., Sarkar, D., Bates, M. D., Matrix, L., 2007. The lme4 package. R package version 2, 74.

Beckman, J., Jessen, M., Ringen, C., 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. Journal of Linguistics 49, 259–284.

Benguerel, A.-P., Bhatia, T. K., 1980. Hindi stop consonants: an acoustic and fiberscopic study. Phonetica 37, 134–148.

Blevins, J., 2004. Evolutionary phonology: The emergence of sound patterns. Cambridge University Press.

Boersma, P., Hamann, S., 2008. The evolution of auditory dispersion in bidirectional constraint grammars. Phonology 25, 217–270.

Boersma, P., Hayes, B., 2001. Empirical tests of the Gradual Learning Algorithm. Linguistic inquiry 32, 45–86.

Boersma, P., et al., 2001. Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341–345.

Bradlow, A. R., 1995. A comparative acoustic study of English and Spanish vowels. The Journal of the Acoustical Society of America 97, 1916–1924.

Bradlow, A. R., Bent, T., 2002. The clear speech effect for non-native listeners. The Journal of the Acoustical Society of America 112, 272–284.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., Bolker, B. M., 2017. Modeling zero-inflated count data with glmmtmb. bioRxiv , 10.1101/ 132753.

Bukmaier, V., Harrington, J., 2016. The articulatory and acoustic characteristics of Polish sibilants and their consequences for diachronic change. Journal of the International Phonetic Association 46, 311–329.

Bukmaier, V., Harrington, J., Reubold, U., Kleber, F., 2014. Synchronic variation in the articulation and the acoustics of the Polish three-way place distinction in sibilants and its implications for diachronic change. In: Fifteenth Annual Conference of the International Speech Communication Association.

Chang, Y.-H., 2013. Variability in cross-dialectal production and perception of contrasting phonemes: the case of the alveolar-retroflex contrast in Beijing and Taiwan Mandarin. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Chang, Y.-H., Shih, C., 2012. Using map tasks to investigate the effect of contrastive focus on the Mandarin alveolar-retroflex contrast. In: Speech Prosody 2012.

Chang, Y.-H. S., Shih, C., 2015. Place contrast enhancement: The case of the alveolar and retroflex sibilant production in two dialects of Mandarin. Journal of Phonetics 50, 52–66.

Chang, Y.-H. S., Shih, C., Allen, J. B., 2013. Dialectal variation in the perception of phonological contrasts. In: Proceedings of the International Conference on Phonetics of the Languages in China.

Chao, Y. R., 1965. A grammar of spoken Chinese. University of California Press.

Chen, F. R., 1980. Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level. Ph.D. thesis, Massachusetts Institute of Technology.

Cheng, C.-C., 2011. A synchronic phonology of Mandarin Chinese, vol. 4. Walter de Gruyter.

Chiu, C., 2009. Acoustic and auditory comparisons of Polish and Taiwanese Mandarin sibilants. Canadian Acoustics 37, 142–143.

Cho, T., Ladefoged, P., 1999. Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics 27, 207–229.

Chodroff, E., Godfrey, J., Khudanpur, S., Wilson, C., 2015. Structured variability in acoustic realization: A corpus study of voice onset time in American English stops. In: Proceedings of the 18th international congress of phonetic sciences. Glasgow, UK: the University of Glasgow.

Chodroff, E., Wilson, C., 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. Journal of Phonetics 61, 30–47.

Chuang, Y.-Y., Fon, J., 2010. The effect of prosodic prominence on the realizations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech. In: Speech Prosody 2010-Fifth International Conference.

Chung, K. S., 2006. Hypercorrection in Taiwan Mandarin. Journal of Asian Pacific Communication 16, 197–214.

Clarke, C., Luce, P., 2005. Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. In: ISCA workshop on Plasticity in Speech Perception.

Clayards, M., 2018. Individual talker and token covariation in the production of multiple cues to stop voicing. Phonetica 75, 1–23.

Clayards, M., Niebuhr, O., Gaskell, M. G., 2015. The time course of auditory and language-specific mechanisms in compensation for sibilant assimilation. Attention, Perception, & Psychophysics 77, 311–328.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., Jacobs, R. A., 2008. Perception of speech reflects optimal use of probabilistic speech cues. Cognition 108, 804–809.

Clements, G. N., 2003. Feature economy in sound systems. Phonology 20, 287–333.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., Wissing, D., 2018. Plosive voicing in afrikaans: Differential cue weighting and tonogenesis. Journal of Phonetics 66, 185–216.

Coolican, H., 2017. Research methods and statistics in psychology. Psychology Press.

Cribari-Neto, F., Zeileis, A., 2010. Beta regression in R. Journal of Statistical Software 34, 1–24. URL `http://www.jstatsoft.org/v34/i02/`.

Cristià, A., McGuire, G. L., Seidl, A., Francis, A. L., 2011. Effects of the distribution of acoustic cues on infants' perception of sibilants. Journal of Phonetics 39, 388–402.

Cyran, E., et al., 2011. Laryngeal realism and laryngeal relativism: Two voicing systems in Polish? Studies in Polish Linguistics 6, 45–80.

Dart, S. N., 1998. Comparing French and English coronal consonant articulation. Journal of Phonetics 26, 71–94.

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

Deterding, D., Nolan, F., 2007. Aspiration and voicing of Chinese and English plosives. In: Proceedings of the 16th International Congress of Phonetic Sciences. Universität des Saarlandes Saarbrücken, Germany.

DiCanio, C., 2013. Time averaging for fricatives 2.0. Praat script published online.

Dixit, R. P., 1989. Glottal gestures in Hindi plosives. Journal of Phonetics 17, 213–237.

Docherty, G. J., 1992. The timing of voicing in British English obstruents, vol. 9. Walter de Gruyter.

Dogil, G., 1990. Hissing and hushing fricatives: A comment on non-anterior spirants in Polish. Manuscript, Stuttgart University .

Duanmu, S., 2000. The phonology of standard Mandarin. Oxford University Press.

Duanmu, S., 2007. The phonology of standard Chinese. Oxford University Press.

Duda, R. O., Hart, P. E., Stork, D. G., 2012. Pattern classification. John Wiley & Sons.

Dutta, I., 2007. Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt. University of Illinois at Urbana-Champaign.

Edmondson, J. A., Esling, J. H., Harris, J. G., Huang, T. C., 2005. A laryngoscopic study of glottal and epiglottal/pharyngeal stop and continuant articulations in Amis - An Austronesian language of Taiwan. Language and Linguistics Taipei 6, 381.

Elston, A. H., Blake, K., Berkson, K., Herd, W., Cariño, J., Nelson, M., Strickler, A., Torrence, D., 2016. Region, gender, and within-category variation in American English voiced stops. The Journal of the Acoustical Society of America 139, 2123–2123.

Esling, J. H., 2003. Glottal and epiglottal stop in Wakashan, Salish and Semitic. In: Proceedings of the 15th International Congress of Phonetic Sciences, vol. 2.

Ferguson, S. H., Kewley-Port, D., 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. Journal of Speech, Language, and Hearing Research 50, 1241–1255.

Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. Journal of Applied Statistics 31, 799–815.

Fisher, R. A., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179–188.

Fougeron, C., Smith, C. L., 1993. French. Journal of the International Phonetic Association 23, 73–76.

Fukunaga, K., 1990. Introduction to statistical pattern recognition. Elsevier.

Garellek, M., White, J., 2015. Phonetics of Tongan stress. Journal of the International Phonetic Association 45, 13–34.

Gendrot, C., Adda-Decker, M., et al., 2007. Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. In: Proceedings of the 16th International Congress of Phonetic Sciences.

Gorman, K., Johnson, D. E., 2013. Quantitative analysis. The Oxford Handbook of Sociolinguistics , 214–240.

Gottfried, T. L., Beddor, P. S., 1988. Perception of temporal and spectral information in French vowels. Language and Speech 31, 57–75.

Greenlee, M., Ohala, J. J., 1980. Phonetically motivated parallels between child phonology and historical sound change. Language Sciences 2, 283–308.

Haeb-Umbach, R., Ney, H., 1992. Linear discriminant analysis for improved large vocabulary continuous speech recognition. In: Proceedings of ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1.

Halle, M., Stevens, K. N., 1997. The postalveolar fricatives of Polish. Speech production and language: In honor of Osamu Fujimura 13, 176–191.

Harnsberger, J. D., Wright, R., Pisoni, D. B., 2008. A new method for eliciting three speaking styles in the laboratory. Speech Communication 50, 323–336.

Harris, J., 1990. Derived phonological contrasts. Studies in the pronunciation of English: A commemorative volume in honour of AC Gimson , 87–105.

Hauser, I., 2017. A revised metric for calculating acoustic dispersion applied to stop inventories. The Journal of the Acoustical Society of America 142, EL500–EL506.

Hazan, V., Simpson, A., 2000. The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. Language and Speech 43, 273–294.

Herd, W., Torrence, D., Carino, J., 2016. Prevoicing differences in Southern English: Gender and ethnicity effects. The Journal of the Acoustical Society of America 139, 2217–2217.

Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K., 1995. Acoustic characteristics of American English vowels. The Journal of the Acoustical Society of America 97, 3099–3111.

Honeybone, P., 2005. Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. The internal organization of phonological segments 319, 54.

Hu, F., 2008. The three sibilants in Standard Chinese. In: Proceedings of the 8th International Seminar on Speech Production.

Hualde, J. I., 2004. Quasi-phonemic contrasts in Spanish. In: Proceedings of the 23rd West Coast Conference on Formal Linguistics, vol. 23. Cascadilla Press.

Hunnicutt, L., Morris, P. A., 2016. Prevoicing and aspiration in Southern American English. University of Pennsylvania Working Papers in Linguistics 22, 24.

Jacewicz, E., Fox, R. A., Lyle, S., 2009. Variation in stop consonant voicing in two regional varieties of American English. Journal of the International Phonetic Association 39, 313–334.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning, vol. 112. Springer.

Jeng, J.-Y., 2006. The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin. Journal of Humanistic Studies 40, 27–48.

Jessen, M., Ringen, C., 2002. Laryngeal features in German. Phonology 19, 189–218.

Johnson, K., Ladefoged, P., Lindau, M., 1993. Individual differences in vowel production. The Journal of the Acoustical Society of America 94, 701–714.

Jongman, A., Fourakis, M., Sereno, J. A., 1989. The acoustic vowel space of Modern Greek and German. Language and Speech 32, 221–248.

Jongman, A., Wade, T., 2007. Acoustic variability and perceptual learning. In: Language experience in second language speech learning: In honor of James Emil Flege. Benjamins Publishing, pp. 135–150.

Jongman, A., Wayland, R., Wong, S., 2000. Acoustic characteristics of English fricatives. The Journal of the Acoustical Society of America 108, 1252–1263.

Kallay, J., Holliday, J., 2012. Using spectral measures to differentiate Mandarin and Korean sibilant fricatives. In: Thirteenth Annual Conference of the International Speech Communication Association.

Keating, P. A., 1984. Phonetic and phonological representation of stop consonant voicing. Language 60, 286–319.

Keshet, J., Sonderegger, M., Knowles, T., 2014. AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [computer program]. version 0.91.

Kim, D., Clayards, M., 2019. Individual differences in the link between perception and production and the mechanisms of phonetic imitation. Language, Cognition and Neuroscience 31, 1–18.

Kingston, J., 1992. The phonetics and phonology of perceptually motivated articulatory covariation. Language and Speech 35, 99–113.

Kingston, J., Diehl, R. L., 1994. Phonetic knowledge. Language 70, 419–454.

Kong, E. J., Yoon, I. H., 2013. L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: Production and perception of English stops. Phonetics and Speech Sciences 5, 81–90.

Krause, J. C., Braida, L. D., 2004. Acoustic properties of naturally produced clear speech at normal speaking rates. The Journal of the Acoustical Society of America 115, 362–378.

Kuang, J., Cui, A., 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. Journal of Phonetics 71, 194–214.

Kubler, C. C., 1985. The influence of Southern Min on the Mandarin of Taiwan. Anthropological Linguistics 27, 156–176.

Kudela, K., 1968. Spectral analysis of Polish fricative consonants. Speech Analysis and Synthesis 1, 93–188.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., 2017. lmertest package: tests in linear mixed effects models. Journal of Statistical Software 82.

Ladefoged, P., Maddieson, I., 1996. The sounds of the world's languages. Oxford: Blackwell.

Ladefoged, P., Wu, Z., 1984. Places of articulation: An investigation of Pekingese fricatives and affricates. Journal of Phonetics 12, 267–278.

Lee, W.-S., 1999. An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin. In: Proceedings of the 14th International Congress of Phonetic Sciences, vol. 413416.

Lee-Kim, S.-I., 2011. Spectral analysis of Mandarin Chinese sibilant fricatives. In: Proceedings of the 17th International Congress of Phonetic Sciences.

Li, F., 2008. The phonetic development of voiceless sibilant fricatives in English, Japanese, and Mandarin Chinese. Ph.D. thesis, Ohio State University, Columbus.

Li, W.-C., 1999. A diachronically-motivated segmental phonology of Mandarin Chinese, vol. 37. Peter Lang Publishing.

Li, Y., 2009. Effects of Lexical Frequency and Neighborhood Density on Incomplete Neutralization in Taiwan Mandarin. Ph.D. thesis, CCU.

Liljencrants, J., Lindblom, B., 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. Language 48, 839–862.

Lin, Y.-H., 2014. Segmental phonology. In: The handbook of Chinese linguistics. John Wiley & Sons, pp. 400–422.

Lindblom, B., 1986. Phonetic universals in vowel systems. In: Experimental Phonology. Academic Press, pp. 13–44.

Lindblom, B., Maddieson, I., 1988. Phonetic universals in consonant systems. In: Language, Speech, and Mind. Routledge, pp. 62–78.

Lisker, L., 1986. "voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. Language and Speech 29, 3–11.

Lisker, L., Abramson, A. S., 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 384–422.

Lisker, L., Abramson, A. S., 1967. Some effects of context on voice onset time in English stops. Language and Speech 10, 1–28.

Livijn, P., 2000. Acoustic distribution of vowels in differently sized inventories–hot spots or adaptive dispersion. Phonetic Experimental Research, Institute of Linguistics, University of Stockholm 11.

Lombardi, L., 1994. Laryngeal features and laryngeal neutralization. Routledge.

Mackie, S., Mielke, J., 2011. Feature economy in natural, random, and synthetic inventories. In: Where do phonological features come from? Cognitive, physical, and developmental bases of distinctive speech categories. John Benjamins, pp. 43–63.

MacNeilage, P. F., 1998. The frame/content theory of evolution of speech production. Behavioral and Brain Sciences 21, 499–511.

Maddieson, I., 1977. Tone loans: a question concerning tone spacing and a method of answering it. UCLA Working Papers on Phonetics: Studies on Tone 36, 49–83.

Maddieson, I., 1981. UPSID: UCLA phonological segment inventory database. Phonetics Laboratory, Department of Linguistics.

Maeda, S., 1990. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: Speech Production and Speech Modeling. Springer, pp. 131–149.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M., 2017. Montreal forced aligner [computer program] version 0.9.0, retrieved from http://montrealcorpustools.github.io/montreal-forced-aligner/.

McCasland, G. P., 1983. Noise segment and vocalic cues of french fricatives. The Journal of the Acoustical Society of America 74, S90–S90.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., 2002. Gradient effects of within-category phonetic variation on lexical access. Cognition 86, B33–B42.

Mielke, J., 2008. The emergence of distinctive features. Oxford University Press.

Mikuteit, S., Reetz, H., 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. Language and Speech 50, 247–277.

Nearey, T. M., 1989. Static, dynamic, and relational properties in vowel perception. The Journal of the Acoustical Society of America 85, 2088–2113.

New, B., Pallier, C., Ferrand, L., Matos, R., 2001. Une base de données lexicales du français contemporain sur internet: Lexique, a lexical database for contemporary French. L'année Psychologique 101, 447–462.

Newman, R. S., Clouse, S. A., Burnham, J. L., 2001. The perceptual consequences of within-talker variability in fricative production. The Journal of the Acoustical Society of America 109, 1181–1196.

Niebuhr, O., Clayards, M., Meunier, C., Lancia, L., 2011. On place assimilation in sibilant sequences: Comparing French and English. Journal of Phonetics 39, 429–451.

Niebuhr, O., Lancia, L., Meunier, C., 2008. On place assimilation in french sibilant sequences. In: 8th International Seminar on Speech Production.

Nittrouer, S., Studdert-Kennedy, M., 1987. The role of coarticulatory effects in the perception of fricatives by children and adults. Journal of Speech, Language, and Hearing Research 30, 319–329.

Nowak, P. M., 2006. The role of vowel transitions and frication noise in the perception of Polish sibilants. Journal of Phonetics 34, 139–152.

Ohala, J. J., 1979. Chairman's introduction to symposium on phonetic universals in phonological systems and their explanations. In: Proceedings of the Ninth International Congress of Phonetic Sciences. Institute of Phonetics, University of Copenhagen.

Ohala, J. J., 1994. Acoustic study of clear speech: A test of the contrastive hypothesis. In: International Symposium on Prosody, vol. 18.

Ohala, J. J., 1995. Clear speech does not exaggerate phonemic contrast. In: Fourth European Conference on Speech Communication and Technology.

Ohala, M., 1983. Aspects of Hindi Phonology, vol. 2. Motilal Banarsidass Publishers.

Padgett, J., Żygis, M., 2007. The evolution of sibilants in Polish and Russian. Journal of Slavic Linguistics 15:2, 291–324.

Picheny, M. A., Durlach, N. I., Braida, L. D., 1986. Speaking clearly for the hard of hearing ii: Acoustic characteristics of clear and conversational speech. Journal of Speech, Language, and Hearing Research 29, 434–446.

Pisoni, D. B., Tash, J., 1974. Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics 15, 285–290.

Poon, P. G., Mateer, C. A., 1985. A study of VOT in Nepali stop consonants. Phonetica 42, 39–47.

Proctor, M., Lu, L. H., Zhu, Y., Goldstein, L., Narayanan, S., et al., 2012. Articulation of Mandarin sibilants: a multi-plane realtime MRI study. In: Proceedings of the 14th Australasian International Conference on Speech Science Technology. Macquarie University, pp. 113–116.

Quirk, R., Greenbaum, S., Leech, G. N., Svartvik, J., et al., 1972. A Grammar of Contemporary English. Longman London.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raphael, L. J., 2005. Acoustic cues to the perception of segmental phonemes. In: The Handbook of Speech Perception. Blackwell, pp. 182–206.

Repp, B. H., Liberman, A. M., Eccardt, T., Pesetsky, D., 1978. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance 4, 621.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., Ripley, M. B., 2013. Package 'mass'. CRAN Repository .

Rose, P., 2010. The effect of correlation on strength of evidence estimates in Forensic Voice Comparison: uni-and multivariate Likelihood Ratio-based discrimination with Australian English vowel acoustics. International Journal of Biometrics 2, 316–329.

Schertz, J., Cho, T., Lotto, A., Warner, N., 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. Journal of Phonetics 52, 183–204.

Schiefer, L., 1986. F0 in the production and perception of breathy stops: Evidence from Hindi. Phonetica 43, 43–69.

Schwartz, J.-L., Boë, L.-J., Badin, P., Sawallis, T. R., 2012. Grounding stop place systems in the perceptuo-motor substance of speech: On the universality of the labial–coronal–velar stop series. Journal of Phonetics 40, 20–36.

Schwartz, J.-L., Boë, L.-J., Vallée, N., Abry, C., 1997. The dispersion-focalization theory of vowel systems. Journal of Phonetics 25, 255–286.

Schwarz, M., Sonderegger, M., Goad, H., 2019. Realization and representation of Nepali laryngeal contrasts: voiced aspirates and laryngeal realism. Journal of Phonetics 73, 113–127.

Scobbie, J. M., 2006. Flexibility in the face of incompatible English VOT systems. In: Laboratory Phonology 8: Varieties of Phonological Competence. Mouton de Gruyter, pp. 367–392.

Shultz, A. A., Francis, A. L., Llanos, F., 2012. Differential cue weighting in perception and production of consonant voicing. The Journal of the Acoustical Society of America 132, EL95–EL101.

Smiljanić, R., Bradlow, A. R., 2005. Production and perception of clear speech in Croatian and English. The Journal of the Acoustical Society of America 118, 1677–1688.

Smith, B. L., Westbury, J. R., 1975. Temporal control of voicing during occlusion in plosives. The Journal of the Acoustical Society of America 57, S71–S71.

Stampe, D., 1972. On the natural history of diphthongs. In: Papers from the Eighth Regional Meeting of the Chicago Linguistic Society. Chicago Linguistic Society, pp. 578–590.

Stevens, K. N., Blumstein, S. E., 1978. Invariant cues for place of articulation in stop consonants. The Journal of the Acoustical Society of America 64, 1358–1368.

Stevens, K. N., Li, Z., Lee, C.-Y., Keyser, S. J., 2004. A note on Mandarin fricatives and enhancement. In: From Traditional Phonology to Modern Speech Processing. Foreign Language Teaching and Research Press, pp. 393–403.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., Nishi, K., 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. The Journal of the Acoustical Society of America 122, 1111–1129.

Summerfield, Q., 1979. Use of visual information for phonetic perception. Phonetica 36, 314–331.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., Stokes, M. A., 1988. Effects of noise on speech production: Acoustic and perceptual analyses. The Journal of the Acoustical Society of America 84, 917–928.

Sussman, H. M., McCaffrey, H. A., Matthews, S. A., 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. The Journal of the Acoustical Society of America 90, 1309–1325.

Tabain, M., 2001. Variability in fricative production and spectra: Implications for the hyper-and hypo-and quantal theories of speech production. Language and speech 44, 57–93.

Theodore, R. M., Miller, J. L., DeSteno, D., 2009. Individual talker differences in voice onset time: Contextual influences. The Journal of the Acoustical Society of America 125, 3974–3982.

Toscano, J. C., McMurray, B., 2010. Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. Cognitive Science 34, 434–464.

Vaughn, C., Baese-Berk, M., Idemaru, K., 2018. Re-examining phonetic variability in native and non-native speech. Phonetica 76, 1–32.

Viszlay, P., Juhár, J., Pleva, M., 2012. Alternative phonetic class definition in linear discriminant analysis of speech. In: 19th International Conference on Systems, Signals and Image Processing. IEEE.

Walley, A. C., Carrell, T. D., 1983. Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. The Journal of the Acoustical Society of America 73, 1011–1022.

Warner, N., Tucker, B. V., 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. The Journal of the Acoustical Society of America 130, 1606–1617.

Well, A. D., Myers, J. L., 2003. Research design & statistical analysis. Psychology Press.

Wolinski, M., Milkowski, M., Ogrodniczuk, M., Przepiórkowski, A., 2012. Polimorf: a (not so) new open morphological dictionary for Polish. In: LREC.

Wright, R., 2004. Factors of lexical competition in vowel articulation. In: Papers in Laboratory Phonology VI. Cambridge University Press Cambridge, pp. 75–87.

Wu, Y., 1994. Mandarin Segmental Phonology. Ph.D. thesis, University of Toronto.

Wu, Z., Lin, M., 1989. Overview of Experimental Phonetics. Higher Education Press, Beijing.

Zeileis, A., Hothorn, T., 2002. Diagnostic checking in regression relationships. R News 2, 7–10. URL `https://CRAN.R-project.org/doc/Rnews/`.

Żygis, M., Hamann, S., 2003. Perceptual and acoustic cues of Polish coronal fricatives. Proceedings of the 15th International Conference of Phonetic Sciences , 395–398.

Żygis, M., Pape, D., Czaplicki, B., 2012. Dynamics of sibilant systems: Standard Polish and its dialects. In: Phonetik & Phonologie 8. Jena: Friedrich-Schiller-Universitat Jena.