EMPIRICAL FOUNDATIONS OF SOCIO-INDEXICAL STRUCTURE:

INQUIRIES IN CORPUS SOCIOPHONETICS AND PERCEPTUAL LEARNING

by

KAYLYNN GUNTER

A DISSERTATION

Presented to the Department of Linguistics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2023

DISSERTATION APPROVAL PAGE

Student: Kaylynn Gunter

Title: Empirical Foundations of Socio-Indexical Structure: Inquiries in Corpus Soociophonetics and Perceptual Learning

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Tyler Kendall	Chairperson
Charlotte Vaughn	Core Member
Vsevolod Kapatsinski	Core Member
Melissa Baese-Berk	Core Member
Kaori Idemaru	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded September 2023

DISSERTATION ABSTRACT

Kaylynn Gunter

Doctor of Philosophy

Department of Linguistics

September 2023

Title: Empirical Foundations of Socio-Indexical Structure: Inquiries in Corpus Sociophonetics and Perceptual Learning

Speech is highly variable and systematic, governed by the internal linguistic system and socio-indexical factors. The systematic relationship of socio-indexical factors and variable phonetic forms, referred to here as *socio-indexical structure*, has been the cornerstone of sociophonetic research over the last several decades. Research has provided mounting evidence that listeners track and exploit cross-talker variability during speech processing tasks. As one such example, recent work has demonstrated listeners' sensitivity to talker characteristics via retuning phonetic categories (i.e., perceptual learning) in response to talker-specific patterns. Drawing on Bayesian models, researchers have argued that listeners' perceptual learning is influenced by listeners' prior experience with socio-indexical factors conditioning segmental variation. From experience listeners form beliefs about the underlying cause of variation to determine when to adapt to talker-specific forms and generalize to other similar talkers. However, theoretical work has over-simplified descriptions of socio-indexical structure, leaving open questions about the nature and range of phonetic variation that listeners track and exploit.

This dissertation seeks to provide both theoretical and empirical foundations of socio-indexical structure at the intersection of individual talkers and geographic dialects drawing on mixed methods. Using large-scale datasets of American English vowel measurements, the corpus analyses probe different quantitative descriptions of socio-indexical structure under various scopes of socio-indexical granularity and internal organizations across the vowel space. The corpus analyses reveal an asymmetry in socio-indexical conditioning of the joint cue distributions (i.e., F1xF2) across several simulations whereby some categories (e.g., /eɪ/) are conditioned by dialect, while others are conditioned by individual talker identity alone (e.g., /ʊ/;

Chapter 4). Additionally, analyses show that individual talkers diverge from their dialect areas less for dialect conditioned vowels compared to talker conditioned vowels, confirming talkers' distributional patterns generally align with their communities. Additional analyses highlight how internal principles provide specificity to socio-indexical conditioning of variability, focusing on the acoustic overlap of vowel pairs and individual cue dimensions (Chapter 5). Such descriptions suggest acoustic overlap across some vowel pairs may be attenuated by socio-indexical information while other vowel pairs generally demonstrate stability across talkers and dialects (e.g., /æ/ and /a/). Finally, descriptions of individual cue dimensions demonstrate multimodal distributions both across and within talkers for some categories conditioned by dialects (e.g., /ɔ/; Chapter 5).

Following from Bayesian models of speech processing and causal inference, this dissertation tests whether a priori links to socio-indexical structure influence perceptual learning (Chapter 6). A lexically guided perceptual learning experiment tests whether the asymmetry of socio-indexical conditioning (dialect vs. talker) observed in the corpus analyses correlates with listeners' learning and generalization behavior after exposure to novel shifts in one of two vowels (/eɪ/ and /ʊ/) in a female speaker's voice. The results demonstrate learning a novel shift in /ʊ/ but not in /eɪ/, with generalization of post-test categorization to a novel male talker but not a novel female talker. These results suggest that the asymmetry of social conditioning alone may guide listeners' behavior for these vowels and challenge our current understanding of listeners' adaptation to vocalic variability and the role of socio-indexical structure in perceptual learning. Overall, this dissertation advances our understanding of socially conditioned variation across speech production and perception.

CURRICULUM VITAE

NAME OF AUTHOR: Kaylynn Gunter

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Nevada, Reno

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2023, University of Oregon
Bachelor of Arts, English, Psychology, French, 2016, University of Nevada, Reno

AREAS OF SPECIAL INTEREST:

Sociophonetics
Sociolinguistics
Psycholinguistics
Corpus phonetics

PROFESSIONAL EXPERIENCE:

Language Data Researcher, Amazon, August 2022 – Current

Graduate Teaching Fellow, University of Oregon, September 2016 – June 2022

GRANTS, AWARDS, AND HONORS:

LSA Linguistic Institute Fellowship, Linguistic Society of America, 2019

PUBLICATIONS:

Gunter, K., & Farrington C. (*in press*). Standards of American English. Wiley Encyclopedia of World Englishes. Wiley Blackwell.

Gunter, K., Vaughn, C., & Kendall, T. (2021). Contextualizing /s/ retraction: Sibilant variation and change in Washington D.C. AAL. *Language Variation and Change,* 33(3), 331-357.

Kendall, T., Vaughn, C., Farrington, C., Gunter, K., McLean, J., Tacata, C., & Arnson, S. (2021). Considering performance in the automated and manual coding of sociolinguistic variables: Lessons from variable (ING). *Frontiers in Artificial Intelligence, 4*.

Gunter, K., Vaughn, C., & Kendall, T. (2020). "Perceiving Southernness: The role of vowel categories and acoustic cues in Southernness ratings." *Journal of the Acoustic Society of America,* 147(1), 643-656.

Gunter, K., Vaughn, C., & Kendall, T. (2018). "Probing the Social Meaning of English Adjective Intensifiers as a Class Lab Project." *American Speech: A Quarterly of Linguistic Usage*, 93(2), 298-311

ACKNOWLEDGMENTS

The number of people who have helped me complete this dissertation demonstrates the most valuable piece of graduate school: community. This process has undoubtedly been challenging for various academic and personal reasons—and without my community I would not have made it to this point. I hope that all of you see how important and valuable you have been throughout this process.

First, I want to thank my committee members Tyler Kendall, Charlotte Vaughn, Kaori Idemaru, Volya Kapatsinski, and Melissa Baese-Berk for their insights, guidance, and flexibility through this process. I also want to thank the administrative team, Eden Cronk, Kayla Robinson, and Em Baker, for making sure I checked all my boxes and letting me swipe candy. Extra thanks to Kayla, whose administrative support turned into friendship and support grew beyond the academic. Thank you to the SPADE team for their collaboration and resources, the undergraduates who participated and provided the voices for stimuli, and Jordan Gallant for help with PsychoPy.

Tyler and Charlotte, thank you for your guidance over the years and helping me grow as a scholar and meeting me where I am but pushing me when I needed it. And thank you both for slogging through this, even when you were tired and busy with your best collaboration, little June. Charlotte, you deserve more thanks than I can express for your mentorship throughout graduate school and this dissertation. I have always considered myself as much your advisee as Tyler's and I am grateful for all that you have done for me despite only advising me in spirit and not in title.

Thank you to my community of (past and present) graduate students and post-docs—you have all made a lasting impact on my research and personal life. I'm grateful to the wonderful students I was lucky enough to teach and learn from over the years, especially Jaidan McLean, Kathryn Paulus, and Carissa Diantoro—it has been a great joy to see you grow as scholars and individuals. Jaidan McLean, thank you for always being willing to help with citations, piloting, stimuli creation, and grabbing coffee. Thank you to the LVC Lab, Jason McLarty, Charlie Farrington, Shelby Arnson, Jaidan McLean, and Chloe Tacata, for listening to many rough and incomplete ideas, presentations, questions, and providing feedback and friendship along the way. I am also grateful for the CogLing group over the years, especially, Eric Pederson and Matt

Stave, for much of the same. Thank you to Zack Jaggers for your friendship and being a great colleague, in academic and non-academic life. Thank you to my cohort Dae-yong Lee and Xuan Guan for being my first friends and officemates at UO. Dae-yong, thank you for the laughter and support through big and small moments over the years, it was a joy to be your "linguistics mom" and friend. Charlie Farrington, thank you for your feedback, sociolinguistic insights, esoteric citations, jokes, The Office quotes, pandemic walks, and friendship. Misaki Kato, thank you for your insight, sympathetic ear, wit, friendship, and sharing an office over the years. Michael McAuliffe, your talents made much of my research possible (MFA, ISCAN, etc.), but your friendship and support contributed so much more; thank you for listening, brainstorming, co-working, sending me pebbles, and letting me hang out with Lapis. Allison and Aaron, thank you both for happy hours and late nights, opening your home to me and Travis for family dinners and holidays, and caring for me. Allison, your scholarship, thoughtfulness, and kindness have always inspired me—I am so grateful to be your friend and collaborator and it has been such a blessing to grow together. Thank you for always celebrating successes and getting me through the challenges.

I also want to thank my family, both given and chosen, for their support over the years and listening to me ramble in joy, frustration, and sadness. Thank you to my mom, Lynn, for always encouraging my love of language and desire to learn; and to my dad, Dave, for always being ready to learn about linguistics, pilot a study, and cheer me on. Thanks to my brothers, Casey and Brandon, and their families for supporting me over the years and sending pictures of my nephews to keep my spirits up. Casey, thank you for the extra time given to listening, encouraging, and making me take a break for some fun. Thank you to the Casalideks—Steve, Cheryl, Capa, and Dan—for showing up with food, wine, and joyful chaos when I needed it. To my closest friends, Isha Patel and Fallon Kimball, for showing up for me, listening to me, encouraging me, and keeping me accountable when I needed it. And finally, Travis, your support and faith in me means more to me than you know. Thank you for showing me love and support at every turn, picking me up when I felt overwhelmed, and always encouraging me to keep going. Not to mention, the many cups of coffee you made me—a necessity for any dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER 1:
INTRODUCTION

Speech is highly variable and systematic. Research over the past several decades in the field of sociophonetics has demonstrated that phonetic variation is systematically governed by internal linguistic and social factors (e.g., Weinreich, Labov, & Herzog 1968), which has since been referred to as *structured variation* (e.g., Chodroff, 2017; Sonderegger et al., 2020; Tanner et al., 2020), which I will use throughout this dissertation. Structured variation is a label given to describe phonetic variation as non-random and systematically organized by internal linguistic principles, and characteristics of talkers and their social groups, the latter referred to as *socio-indexical structure*. Intersecting with this work, research provides mounting evidence that listeners track and exploit talker information to socially evaluate speakers (e.g., Campbell-Kibler, 2011), in processing lexical items (e.g., Goldstone, 1995; Nygaard & Pisoni, 1998; Palmeri et al., 1993), and disambiguating segmental variation (Samuel & Kraljic, 2009; Strand, 1999).

Current theories posit that such behavior results from listeners tracking the statistical contingencies between variability and talkers and groups. As an evidence of this ability, recent work has demonstrated that listeners may perceptually 'retune' their phonetic category boundaries in response to exposure to atypical productions from a novel talker (e.g., /s/ which is more /ʃ/ like perceptually), a phenomenon referred to as *perceptual learning* (Eisner & McQueen, 2005; Norris et al., 2003). Such perceptual learning has illustrated listeners' ability to learn associations of the novel segmental patterns of a talker enabling them to adapt to talker-specific (i.e., idiosyncratic) variation (Reinisch & Holt, 2014; Samuel & Kraljic, 2009; Xie & Meyers, 2017). Additionally, perceptual learning research has demonstrated that these learned patterns may be generalized to other talkers who are acoustically or perceptually similar (Kraljic & Samuel, 2006). Such work has demonstrated sensitivity to prior experience with typological patterns (Babel et al., 2021; Sumner, 2011), contrast type (Eisner et al., 2013; Kraljic & Samuel 2006, 2007; Tamminga et al., 2020), variability in exposure (Sumner, 2011), and cumulative experience with talkers (Lai, 2021; Theodore & Monto, 2019; Tzeng et al., 2020) constraining

perceptual retuning. Current research posits these constraints are informed by listeners' prior beliefs about the distributional properties of contrasts and their relationship with socio-indexical factors from past experience (Kleinschmidt & Jaeger, 2015; Kleinschmidt, 2019; Jaeger & Weatherholtz, 2016).

The growing evidence for the role of socio-indexical factors, in speech production and perception, warrants a comprehensive model of socio-indexical structure of phonetic variation. With this evidence in mind, a comprehensive model must provide an account of how social factors condition phonetic variation in production alongside how listeners learn and exploit this variability in speech processing. Situated in this research context, this dissertation provides initial empirical and theoretical foundations for socio-indexical structure by examining how speech is organized by socio-indexical factors in production and the implications for such structure in perceptual learning. Drawing on theoretical work in sociophonetics and speech processing, I aim to bridge these two bodies of work and lay the groundwork for future advances.

Sociophonetics as a field aims to examine how speech sounds are meaningfully organized both linguistically and socially, drawing from methods in phonetics and sociolinguistic theory (Kendall & Fridland, 2021). While sociophonetic research has grown to include a range of theoretical and methodological interests, the bulk of research in the field has examined regional variation, often through the lens of sound change, and within-individual variation stemming from contextually situated linguistic styles and individual identity (see e.g., Kendall & Fridland, 2021 as evidence). Recent work has also examined the relationship between variable phonetic forms and perception, primarily focusing on *speaker* perception, phonetically cued social inferences about a talkers' social attributes (Kendall et al. 2023). The understanding of socially conditioned variation in *speech* perception has increasingly received attention, primarily focusing on how social information or expectations biases listeners' linguistic categorization (e.g., Campbell-Kibler & miles-hercules, 2021; Drager 2010; McLarty 2019; Niedzielski 1999). However, the role of socio-indexical variation in speech processing tasks, such as perceptual learning, still remains underexamined in sociophonetics despite parallels with sociophonetic interests. Likewise, research in speech processing has underexamined the role of socially conditioned variability in theoretical accounts of how listeners cope with and adapt to phonetic variability. Thus, this dissertation addresses this gap by integrating these bodies of work.

I address the empirical foundations using mixed methods examining vowel variability as a paramount example of socially conditioned variation, focusing on regional dialects and individual talkers as the social factors. This dissertation draws on large-scale corpus phonetics methods as the basis for the measurement and identification of socio-indexical structure in phonetic variation (Chapters 4-5). By using large quantities of naturalistic data, ranging in speaker diversity and speech styles, we are able to model the input which informs listeners' beliefs about how social factors shape speech variability. In contrast, previous computational modeling has drawn on unrealistic distributions from only carefully elicited speech in lab settings (e.g., Kleinschmidt 2019; Kleinschmidt & Jaeger 2015). Thus, this dissertation uses more ecologically representative data to model socio-indexical structure and generate predictions about listener behaviors. Following from this modeling, a lexically guided perceptually learning experiment provides an example of how such corpus techniques can inform experimental work (Chapter 6). Before turning to the primary theoretical background (Chapter 2), I will provide a brief overview of the contents of the dissertation and primary findings.

This dissertation is organized as follows. Chapter 2 provides the theoretical foundation and current gaps in models of socio-indexical structure. Drawing on sociophonetic, psycholinguistic, and phonetic research, this chapter aims to highlight the different perspectives of socio-indexical structure and their limitations. The following three chapters (3-5) relate to the corpus phonetics analyses, the core of the empirical foundations of this dissertation. Chapter 3 provides a general overview of the data and pre-processing for analyses used in Chapters 4-5.

Chapter 4 addresses questions about how diverse experiences with variability shift listeners' a priori assumptions about how socio-indexical structure conditions vocalic variability. This chapter examines and challenges a more generalized perspective of variability whereby socio-indexical factors provide information to listeners about variability across the vowel space, generated in a vowel-specific and multivariate cue space (e.g., F1xF2 for /eɪ/). This perspective suggests listeners' ability to adapt to novel vowel variation occurs symmetrically across vowel categories. To challenge these assumptions, this chapter asks novel questions about how different analytic scopes of social factors (e.g., dialects vs. individual talkers) and variable prior experiences predict alternate listener beliefs. A large barrier to understanding the role of previous experience is that, as of yet, researchers are limited in their ability to measure previous

experience with category variability or socio-indexical factors, especially at scale. Thus, this dissertation attempts to remedy this issue by modeling previous experience through different computational simulations using production data from a large-scale dataset of American English. In terms of social factors, this chapter empirically validates the assumption that individual talkers more regularly and uniformly pattern within social groups. Additionally, this chapter reveals an inverse correlation of the social conditioning of vowel categories, where the acoustic distributions for some vowels (like /eɪ/) are strongly conditioned by dialect information while others are strongly conditioned by individual talker identity (like /ʊ/) across several simulations. In addition, for categories conditioned on dialect information, talkers generally align with their dialect areas more than categories that are conditioned on talkers (but not dialects). In light of these results, Chapter 6 asks to what extent listeners' perceptual learning behavior echoes this asymmetry (see below).

Chapter 5 seeks to nuance the perspectives provided in Chapter 4 and, by extension, previous models of socio-indexical structure in speech processing. As such, this chapter takes a less generalized perspective on how social factors condition variability through analytic approaches more commonly employed in sociophonetics. In this chapter, I draw on a long theoretical tradition in sociophonetics arguing that vowels function as part of a system of interrelated positions. As one aspect of this perspective, I examine how properties of acoustic overlap among vowel pairs are attenuated or not by social factors, providing either flexibility or stability in the vowel space. Additionally, I examine how variability along specific cue dimensions may analogously be structured by social factors, where talkers are more likely to vary along specific cues for certain categories. This chapter illustrates a more fine-grained approach concerning both the socio-indexical factors and more internally governed facets of variability. Given this detail, I hypothesize how such factors may constrain perceptual learning and generalization. While I do not test any of the hypotheses directly, the analysis of these categories is used as part of the criteria for selecting the vowel categories for the experiment in Chapter 6. Additionally, the theoretical and empirical points may elucidate the experimental results, and thus provide critical insights to perceptual learning (as discussed in Chapter 6).

Chapter 6 examines the perceptual learning and generalization of two categories implicated in the analyses in Chapter 4. Here, I hypothesize an asymmetry in learning and

generalization driven by the categories' asymmetry in socio-indexical structure depicted in Chapter 4. Drawing on one facet from Chapter 4, I hypothesized that listeners infer some categories are informative of talkers' dialect background, while others they may infer are idiosyncratic and characteristic of the talker but not a larger dialect area. This stems from the asymmetry observed in Chapter 4, where talkers condition distributional variability in /ʊ/ in informative ways but there is no dialect conditioning of variability. On the other hand, dialects (on average) condition distributional variability in /eɪ/ but not /ʊ/. This asymmetry is hypothesized to lead listeners to a more general updating of /eɪ/ when faced with atypical productions and a more restricted talker-specific updating of /ʊ/. The results, however, do not support this prediction, rather there is learning for the /ʊ/ condition but no learning observed for the /eɪ/ condition and instead a reduction in /eɪ/ responses (increase in /ʊ/ responses). In terms of generalization, both conditions show a generalization of updated category boundaries aligned with post-exposure behavior for a male generalization talker but not a female talker, with both conditions showing an increase in /ʊ/ responses at post-test. This chapter acts as an example of the kinds of questions that naturalistic speech corpora can inform in experimental paradigms, the value of integrating multiple types of data, and the necessity of iteration for refining our theories.

Overall, this dissertation addresses gaps in theories of speech processing by explicating the nature and assumptions of socio-indexical structure. Drawing on methodologies and research across sociophonetics and speech processing allows for a bridging of the respective research areas and deepening insights into the multidimensional nature of socially conditioned variation. As such, it is crucial that we understand the interaction between listeners and talkers through multiple angles and diverse sources.

CHAPTER 2:
BACKGROUND

1 Introduction

Research examining the systematicity of phonetic variation is not new, and indeed there has been a long tradition of identifying the factors that provide order to variation. Weinreich et al. (1968) argued that variation between talkers was constrained by linguistic and social factors, described as 'orderly heterogeneity'. Recent work in (socio)phonetics, has continued to examine the intersection of linguistic form and social factors in the structuring (i.e., organization) of linguistic variability. More recently, researchers have begun to refer to this organization of phonetic forms as 'structured variation' defined as the ways talkers vary from one another in the nature and range of phonetic variation in non-random and systematic (i.e., statistically determined) ways (Chodroff, 2017; Chodroff et al., 2015, Chodroff & Wilson, 2017, 2022; Sonderegger et al., 2020; Tanner et al., 2020). Examining structured variation provides insight into a variety of linguistic theories, including the mapping between phonology and phonetics (Chodroff, 2017), sound change (Sonderegger et al., 2020), and speech perception (Kleinschmidt & Jaeger, 2015). In this dissertation, I am predominately interested in the latter, addressing how socio-indexical structure, a subset of structured variation, aids in listeners' ability to cope with ambiguity and learn novel phonetic variation across talkers.

In terms of speech perception, as most relevant, there is growing evidence that listeners track the statistical contingencies of talkers, social groups, and phonetic forms. As evidence, recent sociophonetic work has demonstrated that listeners categorize talkers into geographic dialect areas (Clopper & Pisoni, 2004c, 2004b, 2004a), illustrating listeners' ability to learn phonetic properties of social groups and infer social identity. Moreover, listeners show fine-grained knowledge about phonetic categories associated with different communities, such that their perception of category boundaries is influenced by the perceived identity of the talker and their own dialect backgrounds (Decker, 2010; Fridland & Kendall, 2012, 2015, 2018; Kendall & Fridland, 2012). Such socially cued perception is thought to be evidence of listeners representing the social contingencies in memory and drawing on them to make social evaluations and guide

predictions for speech categorization (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015). Consequently, much work has theorized that socio-indexical structure is thought to emerge over the distributional properties of individual talkers and social groups (Foulkes, 2010; Foulkes & Hay, 2015; Pierrehumbert, 2003). Furthermore, listeners track the statistical relationship between talker variability and its underlying social causes to leverage during online speech processing tasks such as adapting to novel speech patterns (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016).

This dissertation falls at the intersection of such work, using corpus methodologies to further describe the nature and range of phonetic variability across the vowel space and to lay the groundwork for how such variability is leveraged by listeners during speech processing. I focus specifically on two social factors that condition variability in speech production, geographic dialect areas and individual talker identity. A central purpose of this dissertation is to additionally evaluate the assumptions of approaches to socio-indexical structure in speech processing and to ask what kind of information is necessary to validate listener behavior and observed sociophonetic variability in production. This chapter aims to identify how various levels of analysis (e.g., social, and linguistic) may interact and constrain one another and to further characterize the nature of the problem for speech processing. As such, this background provides an overview of the current perspectives of socio-indexical structure across disciplines in linguistics, primarily sociolinguistics, phonetics, and psycholinguistics. This chapter will first outline what I believe a model of socio-indexical structure must account for, then provide a more extensive overview of what is currently known about each component therein.

2 What is Socio-Indexical Structure?

Socio-indexical structure is defined as the systematic patterns of variable phonetic forms with regard to non-linguistic external social factors. Talkers will vary considerably from each other across a single phonetic cue that is, in part, explained by social factors including broad social groups, such as macro-social categories (e.g., geographic dialect, age, gender), as well as factors including interactional context (Podesva, 2007), emotional state (Kim & Sumner, 2017; Nygaard & Queen, 2008), and so on. Such a definition is undeniably broad, and indeed may reference several related phenomena, but takes as the primary focus patterns of production.

However, as I hope to demonstrate throughout this chapter, a complete representation of socio-indexical structure must bridge what talkers do alongside what listeners know about talkers and groups.

Phonetic variability is constrained by internal linguistic principles of organization such as the correlation between cues in signaling a single contrast (e.g., F1 and F2 among vowels; Clayards, 2018), the covariation (i.e., co-occurrence) among two or more contrasts' phonetic cues (e.g., relationship between /æ/ and /a/ Tamminga, 2019; Kendall & Fridland, 2017), or the correlation of a single cue across contrasts (e.g., VOT in voiced stops; Chodroff & Wilson, 2017). Thus, while socio-indexical structure may refer specifically to the social factors that pattern variability, any comprehensive theory of socio-indexical structure should also address internal principles of phonetic forms, as widely examined by sociolinguists. Internal principles encompass a range of phenomena, of which this dissertation will only examine some limited subset primarily revolving around phonological principles, motivated by questions in speech processing and a large body of work in sociophonetics (described in Section 7 below and Chapter 5).

As noted above, I seek to bridge the relationship between socio-indexical structure as evident in speech production, to how listeners take advantage of socio-indexical structure to guide perceptual behavior, focusing on the perceptual retuning of phonetic categories, henceforth referred to as perceptual learning. Socio-indexical structure, as theorized in some models of speech processing, has been largely bifurcated into two social dimensions that predict between-talker phonetic variability: social groups (e.g., dialect or gender) and individual talkers (Kleinschmidt, 2019). Under these social dimensions, socio-indexical structure is theorized to be emergent in two key ways: 1) the statistical relationship between phonetic variability and social factors in speech production, and 2) listeners' representations of socially cued phonetic variation are learned through experience with talkers. To this end, a talker or group's 'accent' has been formalized as a cue distribution of a given contrast (e.g., VOT of stops for Talker A, and of Dialect A; Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). Talkers are predicted to be consistent in how they produce cues to categories and multiple talkers cluster into groups such as regional background and gender, as a consequence of such consistency (Kleinschmidt, 2019). Thus, language users form bottom-up representations of

structured variation as a consequence of interactions with similar talkers over the course of exposure to tokens across talkers. Consequently, listeners form beliefs about categories informed by their prior experience and make probabilistic inferences to guide subsequent behavior in perceptual learning.

In light of this, this dissertation focuses on the distributional properties of phonetic cues for vowel categories and their systematic relationship with social factors (regional dialect and individual talkers) in American English. While the findings and questions outlined in this dissertation are relevant to a range of questions in phonetics, phonology, and sociolinguistics, the primary discussion will focus on how socio-indexical structure in production is related to processes of what listeners do in perceptual learning. This work provides an underexamined description of the vowel space in terms of the variability of contrasts within and across talkers. I will argue that by looking at the distributional properties of the vowel space, as most exemplifying socio-indexical structure in American English, we are better able to unify related phenomena across otherwise disparate literatures. Quantifying variability across the vowel space has important implications for sociophonetic theory, thus additionally filling a necessary methodological gap for the field. Similarly, turning to sociophonetics can provide valuable insights for speech processing, including the description of socially meaningful variation that represents the input from which listeners learn.

With these goals in mind, I will aim to refine the theoretical scope and delineate what I believe to be the critical components of linguistic and social behavior a model of socio-indexical structure must account for, including:

- Group conditioned variability of phonetic contrasts (Section 7.2.1); How are social groups defined? How does such a delineation predict phonetic variability?
- Individual variability of phonetic contrasts (Section 7.2.2); What is the statistical relationship between individual talkers and their social groups? How similar are talkers within a given social group?
- How do phonetic dependencies of contrasts and individual cue dimensions (i.e., internal factors) constrain or interact with socio-indexical structure? (Section 7.2.2 & Chapter 5)

- How does what talkers do map onto what listeners know and do with the speech signal? (Section 4)
- How do different experiences and conceptualizations of variability shape people's knowledge of socio-indexical structure and their inferences about phonetic variability? (Section 5.2, 6)

I will review each of these factors in-depth and outline what they mean for a model of socio-indexical structure in the proceeding sections.

Before moving into the theoretical foundations, it's important to define terms often used to describe variability across linguistics: variation, variability, and variance. First, variation here refers to contextual differences in the realization of linguistic variants, whereby contextual factors may refer to linguistic (e.g., phonological context) and social (e.g., regional dialect; Labov, 1969; Fasold, 1991) elements. Historically, the use of impressionistic coding of categorical variants (i.e., presence or absence) and probabilistic conditioning of those discrete variants across groups (Labov, 1969, 1972) were the primary means of capturing variation in language use by sociolinguistics. Such a distinction has primarily resulted in the examination of central tendency (e.g., means) to characterize the speech of groups. In contrast, variability is often used more broadly to reference stochastic fluctuations inherent of a continuous distribution of phonetic cues, which may arise from both meaningful differences across talkers (e.g., vocal tract length) and some degree of randomness in the speech signal or from measurement error. As a result, an examination of variability may frequently reference the spread and properties around a central tendency. Finally, both notions can be separated from variance, which is a statistical measure of variability, capturing the spread and deviation of data from the mean (see also Vaughn, Baese-Berk, Idemaru, 2019 for similar definitions). The distinction between *variation* and *variability* across sub-domains has resulted in disparate bodies of work and incomplete integration when overlap does exist. However, for a complete perspective of 'talker variability,' we must examine both the differences between groups' averages and also the properties of variability around the mean.

3 Vowels & Socio-Indexical Structure

Vowels have been the focus of much work in sociophonetics as a primary locus of socially meaningful variation in English (Labov 1994, 2001; Thomas 2011) and much attention has been devoted to describing the internal organization of vowels (e.g., Weinreich et al., 1968; Labov et al., 1972; Labov, 2001; Thomas, 2001). Much of this work has examined the role of variation in processes of language change and the properties of vocalic shifts (Eckert, 1980, 1988; Labov, 1994, 2001, 2010; Labov, Yaeger, & Steiner, 1972; Labov et al., 2006) emphasizing that vocalic variation operates within a system of related changes (i.e., chain shifts) rather than as individual category changes (Labov, 1994, 2001; Labov et al., 2006). While the focus of this dissertation examines variability across vowels as static synchronic systems, the insights from research on vowel shifts provide critical insight into the social and phonological organization of language variation. In addition, since the social factor of primary interest in this dissertation is regional dialect, it's important to understand the patterns of vocalic variation associated with regions in the U.S., as well as the regularities of behaviors associated with vowel shifts more broadly. Thus, vowels provide a paramount example of the social and phonological systematicity of phonetic variation, making them a good case study for understanding socio-indexical structure more broadly. Below I will overview the major patterns of regional dialects before moving on to the broader theoretical discussion. Regional variation of vowel shifts critically provides expectations for analyses in Chapters 4 and 5, with more specific predictions and details within the respective chapters.

3.1 Contextualizing Regional Vowel Shifts

3.1.1 Northern Cities Shift (NCS)

The Northern Cities Shift (NCS) is a chain shift described as involving several related shifts in the vowel space, including the raising, fronting, and diphthongization of /æ/, the fronting of /a/, the lowering of /ɔ/, the backing of /ʌ/, and the retraction of /ɛ/ and /ɪ/ (Eckert, 1989; Nesbitt, 2018; Durian & Cameron, 2018; D'Onofrio & Benheim 2020; Labov et al., 1972). As alluded to, the NCS also maintains a distinction between the low back vowels, in contrast to some other regions of the U.S. The chain shift is schematized in Figure 2.1, presenting the

canonical position alongside the direction of the shift, end points of the arrows represent expectations of the static outcomes of the shift, abstractly. Figure 2.4, adapted from Labov et al. (2006), shows the geographical area where the NCS occurs (labeled as the North). In terms of socio-indexical structure, we can broadly expect that talkers within the Northern U.S. should regularly group together in the conditioning of the phonetic distributions of each of the individual vowels implicated in this shift. Additionally, we can predict that multiple vowels may pattern together, such that if talkers in the North have a raised and fronted /æ/ the likelihood of having a fronted /a/ is higher (e.g., Tamminga 2019); I will come back to this point in more detail in Chapter 5.

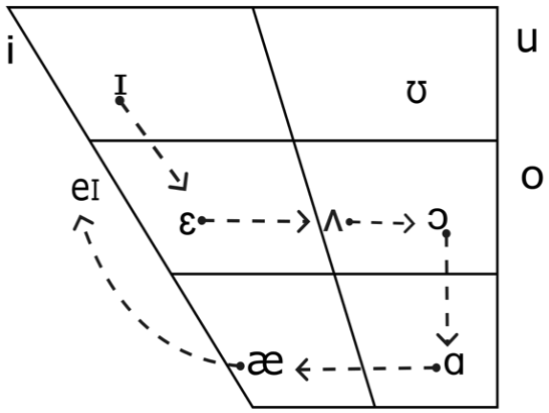

Figure 2.1 Northern Cities Shift schematic, illustrating the canonical English vowel positions and direction of shifts. Note, diphthongs are not included here for ease of reading.

3.1.2 Southern Vowel Shift (SVS)

The Southern Vowel shift is described predominately by the movement of the front tense and lax vowels, whereby the lax vowels move towards the periphery of the vowel space and the tense vowels move towards the center of the vowel space, reversing in phonetic space (Fridland, 2000; Fridland & Kendall, 2015; Labov et al., 2006). The SVS is also characterized by /aɪ/ monophthongization (Feagin, 1986; Fridland, 2000, 2003, 2012; Fridland & Kendall, 2015), and /æ/ raising and diphthongization (Sledd, 1966; Feagin, 1986; Thomas, 2003; Koops, 2014), or in more extreme cases triphthongization (Feagin, 1986). In addition to these core components of the SVS, Southern speech is also characterized by distinct low-back vowels (Fridland & Kendall,

2015; Kendall & Fridland, 2017; Labov et al., 2006; Thomas, 2001) with upgliding of /ɔ/ (Irons, 2007; Thomas, 2001) and fronting of the tense back vowels /o/ and /u/ (Fridland, 2000; Labov et al., 2006). While originally thought to be a hallmark of Southern speech, back vowel fronting has become more prevalent across the U.S. but remains most advanced in the South (Fridland & Bartlett, 2006). Here again, Figure 2.2 presents a schematic of the vowel shifts characterizing the SVS and Figure 2.4 shows the broad geographical region where the SVS occurs (labeled as the South).

While extreme versions of the SVS show a complete reversal of the high and mid front vowels, the categories typically become more proximal in phonetic space, with decreased distance between category means (Fridland, 2000; Fridland & Kendall, 2015). High front vowel shifts, however, are typically restricted to the deep South and other sub-regions generally maintain more canonical positions for these categories. Contrastingly, there is considerable evidence that the mid-front vowels are critical in the SVS and are reflective of a more widespread vowel system across the South (Fridland, 2000; Labov et al., 2006; Kendall & Fridland, 2012). In terms of socio-indexical structure, then, we would expect that the Southern regional affiliation should condition phonetic variation across many vowels but may be most evident for the mid-front vowels (see e.g., Fridland & Kendall 2012) and greater individual variability for high front vowels.

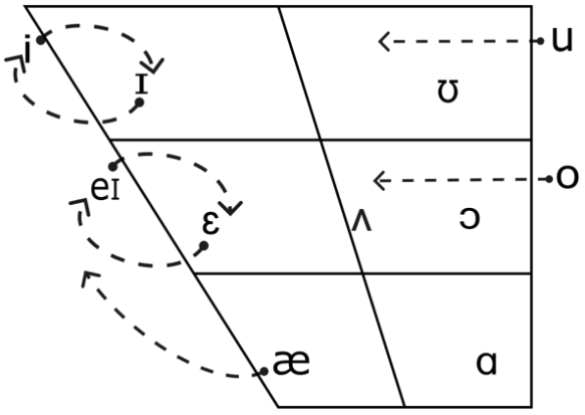

Figure 2.2 Southern Vowel Shift (SVS) schematic, illustrating the canonical English vowel positions and direction of shifts. Note, diphthongs are not included here for ease of reading.

3.1.3 Low-Back Merger Shift (LBMS)

The Low-Back Merger Shift (LBMS) also called the California Vowel Shift (CVS; Hall-Lew, 2009; Eckert, 2008; Podesva et al., 2015), Canadian Vowel Shift (Clarke et al., 1995; Boberg, 2005), and the Elsewhere shift (Fridland, Kendall & Farrington, 2013; Stanley, 2020), is a vowel shift that affects much of the U.S. (as indicated by the name "Elsewhere shift"), but is largely associated with speech in the Western U.S. and Canada. I will use the Low-Back Merger Shift (LBMS; Becker, 2019; Fridland & Kendall, 2019) to refer to it for the entirety of this dissertation. This vowel shift is marked by the low-back vowel merger, as the name suggests, and the lowering and/or retraction of front lax vowels and fronting of the tense back vowels /u/ and /o/. Again, Figure 2.3 provides a schematic of the vowel shift and Figure 2.4 shows the broad dialect area where the shift occurs (labeled as the West).

Several scholars have suggested the low-back vowel merger is the triggering event for shifts in the front lax vowels, (Bigham, 2010; Kendall & Fridland, 2017; Grama & Kennedy, 2019; Labov et al., 2006), hence the nomenclature. Recent work has examined whether there is a structural relationship between the merger of the low back vowels and /æ/ position and has shown that across major geographic regional shifts, the relative distance between /æ/ and /a/ remains similar, despite shifts in the mean positions of each category (e.g., Kendall & Fridland 2017). Therefore, the Western U.S. dialect area conditions phonetic variation across much of the vowel space but /æ/ and /a/ are relationally unique in the LBMS.

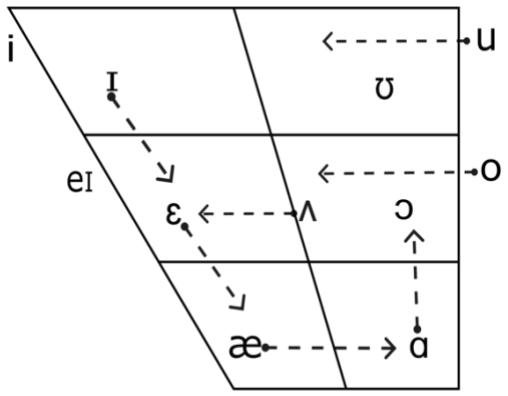

Figure 2.3 Low-Back Merger Shift (LBMS) schematic, illustrating the canonical English vowel positions and direction of shifts. Note, diphthongs are not included here for ease of reading.

Overall, vowel shifts have offered important insights into how social structure is essential to the function of the linguistic system and demonstrates orderly heterogeneity (Eckert, 2000; Weinreich et al., 1968; Labov et al., 1972; Labov et al., 2006, among others). The vowel shifts outlined here have been the focus of much sociophonetic work in the last several decades and have aided in our understanding of how social factors predict linguistic variation. In addition to these larger patterns of sound change, sociophonetics has offered extensive insight into the socio-indexical factors that promote variability within larger regional shifts and the social meaning of individual elements embedded in styles and has provided essential methodological insights into advancing the study of the vocalic system.

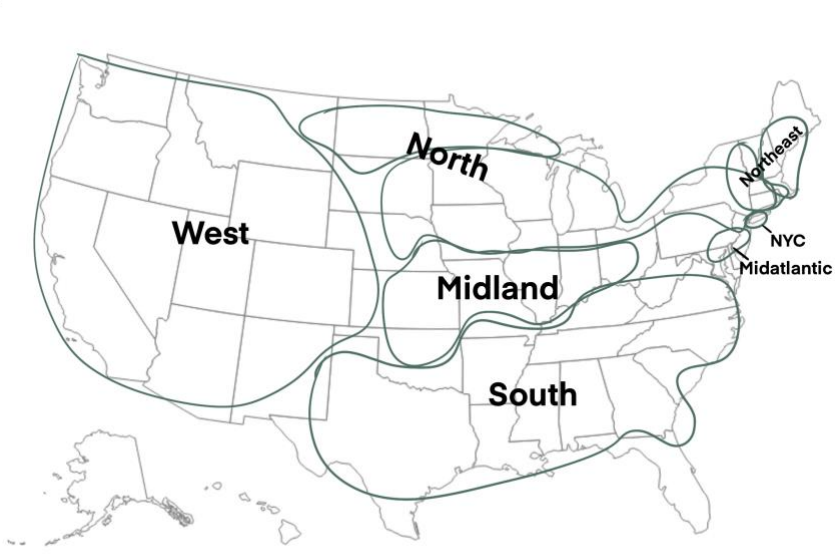

Figure 2.4 Broad regional isoglosses drawn in blue, adapted from Labov et al., 2006. Regional areas are labeled, broadly, with the West representing the LBMS, the North the NCS, and the South the SVS.

4 Socio-Indexical Structure in Perception

Regional vowel shifts garnered much attention over the last several decades and have prompted a more comprehensive understanding of listeners' perceptual evaluations and inferences concerning such phonetic forms. Overall, research reifies regional dialects as a meaningful social group for listeners, as evidenced by shared latent knowledge about regional

patterns of phonetic variability, despite variable prior experience with dialects. Much of the work on sociolinguistic perception has focused on the attributes that listeners associate with different variable forms and perceived talker identity, which I will refer to as *speaker* perception (following Kendall et al. 2023). Additional evidence can be found in *speech* perception work examining how social information influences linguistic comprehension and categorization. Taken together, both speaker and speech perception provide evidence of learned patterns of variable phenomena that listeners leverage for a variety of tasks (see e.g., Campbell-Kibler, 2010 and Thomas, 2002 for a review). Below, I limit the scope of review to a subset of literature in sociophonetics that illustrates listeners' ability to categorize talkers by regional dialect and calibrate their linguistic perception (i.e., speech perception) in accordance with regional variation. Following this review, I will turn to some of the current theoretical perspectives on how listeners learn and represent such socially conditioned variation for speech perception.

Research on speech perception has demonstrated that listeners retain fine-grained phonetic knowledge of socially meaningful variation and expectations about talkers can affect speech perception (Babel, 2009; Staum Casasanto, 2010; Strand & Johnson, 1996; Szakay, Babel, & King, 2016; Vaughn, 2019; Walker & Hay, 2011). Listeners make use of talker identity to guide their inferences in speech processing such as for lexical retrieval (Babel, 2009; Hay & Drager, 2010; Staum Casasanto, 2010; King & Sumner, 2014; King & Sumner, 2015; Sumner & Kataoka, 2013), and for the categorization of speech sounds (Strand, 2000; Johnson, 2006; Kendall & Fridland, 2012, 2017). Some work has also demonstrated that the processing of speech may be influenced by expectations of the talker, where processing is impaired when expectations of social and acoustic cues misalign (Koops et al., 2008; Rubin, 1992; Vaughn, 2019) or enhanced when expectations align (McGowan, 2011; Szakay et al., 2016).

Several studies have demonstrated that previous experience with dialects aids speech processing (Evans & Iverson, 2007; Floccia et al., 2009; Scott & Cutler, 1984; Impe et al., 2008; Adank et al., 2009). Listeners demonstrate an asymmetry in processing speech where exposure to standardized varieties facilitates processing but exposure to marked varieties impairs processing (Floccia et al., 2009; Scott & Cutler, 1984; Adank et al., 2009). The processing gain of standardized varieties has been argued to be the result of standardized talkers, as the perceivers, having less exposure to the marked variety, while the reverse is not as likely. Improvements in

processing as listeners gain exposure to marked variation outside of their own dialect further supports this interpretation (Sumner & Samuel, 2009; Impe et al., 2008). For example, Sumner and Samuel (2009) found that listeners in New York were primed by non-rhotic variants before semantically related target words (e.g., slend[ə] primes thin), regardless of whether they produced non-rhotic forms in their own speech or were raised in the area. Listeners who lacked experience with non-rhotic dialects, on the other hand, showed no evidence of priming effects for the non-rhotic variants. This study illustrates that listeners' previous experience with dialect areas may facilitate speech processing but that listeners can overcome such effects through exposure and, critically, continue to accumulate experience with variability over their lifetime.

In the realm of speaker perception, a considerable amount of work has examined listeners' ability to identify and categorize regional dialects, often correlating the phonetic cues that may aid in listeners' ability to effectively categorize talkers. Much of this work has demonstrated that listeners are sensitive to variants of regional vowel shifts when identifying talker dialect, such as elements of the Southern Vowel Shift (Fridland et al., 2004; Plichta & Preston, 2005). In particular, Clopper and Pisoni (2004b, 2004c, 2007) found that listeners were able to categorize talkers into dialect areas in the U.S. above chance and performed best when grouping talkers into 4-6 regions, over more fine-grained divisions among dialect areas. Confusions in categorization were almost entirely the result of more similar dialect areas being grouped together, and unsurprisingly more dissimilar dialects were less likely to be confused (Bradlow et al., 2010; Clopper et al., 2006). In such cases, it's clear that listeners have some a priori knowledge about the links between variable forms and the regional identity of talkers and can generalize such patterns across talkers of the same regional background.

In another study, Clopper and Pisoni (2004c) demonstrate that listeners used some phonetic cues previously illustrated in studies of dialect variation; however, even when some of the most salient cues were absent, listeners still correctly categorized talkers from other available acoustic cues. Clopper and Pisoni (2004b) further showed that early linguistic experience facilitated listeners' ability to correctly categorize talkers into dialect areas, such that those who were highly mobile earlier in development ("army brats") were more accurate in dialect categorization than listeners who had a more sheltered and less mobile life. Performance gaps can be attenuated when listeners are provided adequate training and feedback during dialect

classification, demonstrating learning of the broad perceptual characteristics of dialect areas (Clopper & Pisoni, 2004a). The work by Clopper and Pisoni (2004a, 2004b, 2004c) illustrates listeners' sensitivity to fine phonetic detail and highlights low-level knowledge of talker and group characteristics for inferences about geographic dialects. Furthermore, while listener sensitivity may vary as a function of previous experience and mobility, listeners nonetheless develop shared latent knowledge of geographically based dialect variation in the U.S.

In a similar vein, Gunter et al. (2020) suggest that listeners track details of talker-specific phonetic variation to make evaluations of regional accentedness. In this study, listeners participated in an accent rating task which exposed them to talkers from the Southern U.S. who varied in their participation in the SVS. As indicated in Section 3.1.2, the movement in the mid-front vowels has been identified as a critical aspect of the SVS, whereby talkers who are shifted may demonstrate these vowels becoming more proximal in acoustic space or reversing altogether. Gunter et al. (2020) demonstrate that the distance between individuals' /eɪ/ and /ɛ/ categories predicts accent ratings of the talkers, with more proximal /eɪ/-/ɛ/ resulting in higher accent ratings across items. Additionally, the results illustrated that listeners rated individual vowel categories as higher in Southern accent ratings that comprised more statistically regular phonetic cues to the group, such as /u/ position and the low back vowels where all talkers in the region tended to share the same acoustic positions, despite individual differences among the more salient categories (e.g., /eɪ/ and /ɛ/). Such a finding illustrates that listeners were able to use prior experience with the dialect area to make evaluations and learn about the properties of individual talkers in the experiment to make evaluations of regional identity. Moreover, they argue listeners track the phonetic dependencies between categories (/eɪ/-/ɛ/) in addition to individual categories' cue distributions (e.g., /u/) for such evaluations. Overall, this experiment suggests listeners are able to learn structured variation and the relationships between individual talkers, their social groups, category cue distributions, and dependencies between categories to evaluate regional accents.

Several studies have also examined how listeners' perceptual category boundaries may shift depending on the vocalic patterns that typify their own dialect. In particular, in a series of studies conducted by Kendall and Fridland (Fridland & Kendall, 2012; Kendall & Fridland, 2012, 2017) listeners' dialect backgrounds and production patterns were found to predict their

categorization boundaries along pairs of vowels related to different ongoing shifts across the U.S. In one such example, listeners who participated in the SVS as evidenced by more proximal mid-front vowels perceived the boundary between /eɪ/ and /ɛ/ differently than listeners from other regions (Fridland & Kendall, 2012). A similar pattern was observed for the low vowels /æ/ and /a/ which are considered critical to several regional shifts and related to low back vowel merger (Bigham, 2010; Gordon, 2005; Kendall & Fridland, 2017; Thomas, 2001). Kendall and Fridland (2017) examined the relationship between talkers' /æ/ and /a/ categories across regional varieties and found that across dialect groups, while the overall positions in the vowel space varied by group, the distance between the categories was maintained across dialect areas. In addition, they found that talkers' boundary between /æ/ and /a/ was predicted by their degree of low-back vowel merger in production but not their regional affiliation or individuals' production of /æ/ and /a/. Such a finding illustrates that vowel perception may be guided by the more systemic relationships among vowels.

Research has also demonstrated that listeners shift their categorization boundaries of vowels based on the perceived characteristics of the talker. A series of studies by Plichta and Rakerd (Plichta & Rakerd, 2010; Rakerd & Plichta, 2010) found that Detroit listeners' boundaries of /a/ and /æ/ shifted depending on whether they were listening to another person from Detroit or the Upper-Peninsula of Michigan. The tendency to shift phonetic percepts has also been demonstrated when only top-down information signifying the talker's regional identity is presented. For example, listeners in Detroit report hearing more Canadian Raising when they are told the talker is from Canada than when they are told the talker is from Detroit (Niedzielski, 1999). This effect has been replicated in New Zealand, where listeners report percepts that are associated with either Australia or New Zealand based on the talker's labeled linguistic background (Hay et al., 2006) or by the mere presence of a stuffed Kiwi prompting nationality inferences (Hay & Drager, 2010). Taken together, such findings illustrate that listeners have expectations about the boundaries between vowel categories and relationships among categories based on their own system and experience with dialect areas (their own or others; see Chapter 5 for additional discussion).

Such perceptual flexibility has been suggested to prompt changes in individuals' production behavior. Some work has demonstrated this relationship in the case of second dialect

acquisition, where talkers not only update perceptual representations but may later shift their speech more in line with the new community they belong to (Nycz, 2015). In addition, listeners have been shown to reproduce sociophonetic variation in response to their interlocutors' speech, which may further be mediated by social perceptions. Listeners exhibit greater convergence towards phonetic variants when they positively evaluate their interlocutor (Babel, 2012; Natale, 1975) and, on the other hand, greater divergence when social distance is greater (Bourhis & Giles, 1977; Abrego-Collier et al., 2011; Yu et al., 2013). Such accommodation has been demonstrated to occur based on inferences about the talkers' regional background and expectations of their speech even in the absence of direct evidence. In particular, Wade (2022) demonstrates that listeners converge towards unheard vocalic variants of a Southern U.S. talker, signifying inferences about the relationship of vowel categories in regional dialects. In this experiment, listeners converged towards monophthongal /aɪ/ after exposure to a talker with Southern-shifted speech, despite never hearing the model talker produce the vowel category.

Accommodation has been demonstrated for non-salient regional variation as well. Specifically, research has shown that talkers with vowel mergers may unmerge in response to their interlocutor's system (Babel et al., 2013; Hay et al., 2009). This finding has been used to suggest that near-mergers, whereby talkers may be merged in production but not in perception, may be the result of tracking talker variability in their broader communities (Hay et al., 2009). This work in accommodation further provides evidence that listeners track the statistical regularities of social and linguistic forms to exploit in online communication, even in cases where the variability is not socially salient, as in vowel mergers (Labov, 1994; Eckert & Labov, 2017). Research on accommodation has illustrated that listeners track the statistical contingencies between talkers and phonetic variation and build expectations about these contingencies.

Work examining speaker and speech perception provides evidence that listeners are sensitive to subtle acoustic variation in speech and prior experience with such variation influences speech processing. Such evidence points to the fact that listeners must learn and represent socio-indexical structure for exploitation during a variety of tasks. Listener-oriented behaviors must be learned from previous experience with language, and the only clear place is from sampling the noisy acoustic distributions in their environment from experience with talkers.

Such a fact has prompted the integration of socio-indexical structure into usage-based frameworks of learning and representation.

5 Representation of Socio-Indexical Structure & Phonology

Usage-based theories have provided a theoretical framework through which socio-indexical structure can feasibly be incorporated into cognitive representations. Much of the current work in this area draws on exemplar theoretic frameworks (Docherty & Foulkes, 2014; Foulkes & Docherty, 2006; Pierrehumbert, 2001, 2006; Sumner et al., 2014) due to their robust ability to cogently connect the social facts of linguistic variation, observed by sociolinguists over the last several decades, into cognition. Socio-indexical knowledge is thought to be an emergent property over experienced distributions in phonetic space via social interaction (Docherty & Foulkes, 2014; Foulkes & Docherty, 2006; Foulkes & Hay, 2015; Foulkes, 2010; Kleinschmidt, 2019; Pierrehumbert, 2003, 2006).

The guiding principle of exemplar theory is that language users encode detailed instances of spoken words in episodic memory, and, in processing speech, representations of lexical items are activated as a function of the acoustic similarity to the incoming speech signal (Johnson, 1997; Pierrehumbert, 2001, 2003, 2016; Sumner et al., 2014). Instantiations of exemplar frameworks may vary in exactly how much abstraction occurs over experienced exemplars. Some scholars argue representations are based on lexical items and experienced exemplars (Johnson, 1997) while others call for more hybrid models allowing for abstraction to occur over exemplars to sub-lexical linguistic units (e.g., Norris & McQueen, 2008; Pierrehumbert, 2003, 2006, 2016). Instantiations of exemplar models further describe the encoding of the socio-indexical information that is linked to the acoustic representation, such as talker identity (Goldinger, 1997), gender (Johnson, 2006), and speech context (Local, 2003). Critically, such work explains the observations of speaker and speech perception in Section 4, as listeners can map the incoming speech signal to representations of acoustically similar events and their shared associations of socio-indexical information.

Much work has demonstrated that listeners learn talker-specific details, enabling the ability to distinguish individual talkers from one another. At some stage, listeners may generalize experiences with multiple talkers to broader groups of talkers (Foulkes & Docherty, 2006;

Foulkes & Hay, 2015; Pierrehumbert, 2003). Evidence of both talker and group knowledge is illustrated by known talkers facilitating lexical access (Bradlow et al., 1999; Nygaard & Pisoni, 1998) and voices with more stereotypical gendered voices demonstrating improved processing over atypical individuals within their gender group (Strand & Johnson, 1996; Strand, 1999, 2000). Thus, exemplar theoretic frameworks allow for talker-specific encoding speech patterns alongside more generalized links to social factors. Such encoding is thought to occur over acoustic space, as talker identity and social factors may functionally partition the variable acoustic space.

Variability conditioned on individual talker identity and gender is thought to be more discretely and robustly partitioned in phonetic space compared to regional dialects, as is evidenced by large differences in F0 between genders resulting from physiological and social factors (Foulkes & Hay, 2015). Such relationships are thought to be more readily emergent and easier to learn, while the more arbitrary associations between social groups and variation may take longer to learn and knowledge may emerge later (Foulkes & Hay, 2015; Foulkes & Docherty, 2006). Such a theory makes the examples in Section 4 more challenging to resolve, as regional identity may be both increasingly arbitrary and encompass a much noisier set of exemplars from which associations emerge. Yet, regardless, American English listeners have some degree of shared knowledge about regional variation and are able to learn the relationships between talkers and regional groups with relatively little exposure[1], such as over the course of an experiment (Clopper & Pisoni, 2004a; Gunter et al., 2020). Furthermore, such knowledge influences how listeners approach phonetic categorization (Fridland & Kendall, 2018; Hay et al., 2006; Hay & Drager, 2010), illustrating the necessity to examine variability and the associated socio-indexical relationships' impact on phonetic categorization.

5.1 Distributional Learning

An underlying mechanism enabling language users to learn rich details of socio-indexical information may be distributional learning, where listeners acquire knowledge about how often

[1] Sumner et al. (2014) incorporate a dual-route weighting mechanism where low frequency and highly salient items may be weighed more heavily. Given the emphasis of this dissertation is around raw frequency distributions and resolving phonetic category ambiguity, I leave such an integration for future work.

various kinds of stimuli occur in the environment[2]. Distributional learning refers to learning category structure from frequency distributions, most often evidenced by the learning of sound categories. As an example, if given two sound categories, /p/ and /b/, the Voice Onset Time (VOT) distribution (Figure 2.5) should have a bimodal frequency distribution in the environment as the cue critically distinguishes the categories separated by different central tendencies (means or modes). The statistical input provides critical information for bootstrapping novel sound categories by partitioning the continuous distribution into discrete phonetic categories (Maye et al., 2002; Maye et al., 2008; Quam & Creel, 2021).

Indeed, such learning has been demonstrated in first language acquisition where children learn the stochastic properties of ambient speech sounds (Saffran et al., 1996; Maye, et al., 2002; Maye et al., 2008; Quam & Creel, 2021) and adults when learning sound contrasts in a second language (Baese-Berk, 2010; Wright et al., 2015). Listeners may also learn to shift attention to a single cue dimension from multidimensional contrasts even when the cue relationships are inverted from expectations of typical productions (Harmon et al., 2019; Kruschke, 1996). The growing body of evidence points to the fact that listeners learn acoustic cue distributions utilizing the variability as a cue to contrastive speech sounds and maintain malleability of such contrasts over their lifetime.

[2] Kapatsinski (2018) suggests that distributional learning may not be a unique learning mechanism, but rather the outcome of several learning mechanisms working together.

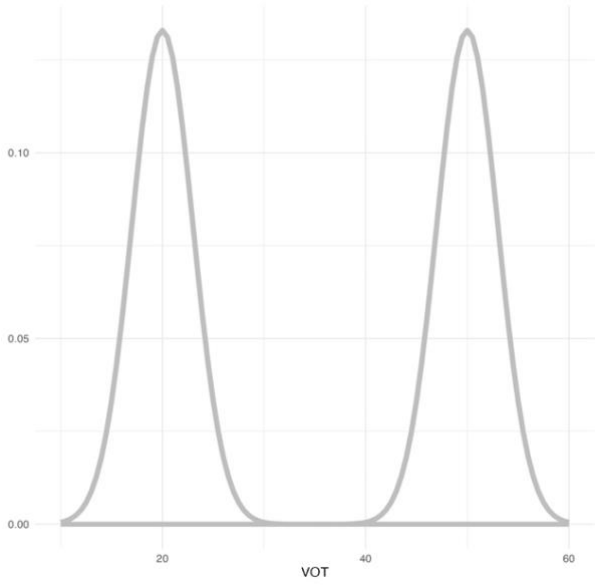

Figure 2.5 Example of bimodal distribution of VOT where the /b/ category has an average VOT of 20ms and the /p/ category has a VOT of 50ms. Such a distinction might show category differences by mean VOT and share the same variance.

In addition, research has suggested that listeners are sensitive to the parameters of individual sound categories after learning has occurred. Cohen et al. (2001), while not examining sound categories, demonstrate that if listeners are given a stimulus equidistant from two category boundaries, listeners will assign the token to the category that has a wider range of within-category variability compared to the narrower variability category. This effect is demonstrated even when the stimulus item is closer to the mean of the narrower distribution than the wider distribution. As a depiction, Figure 2.6 shows hypothetical variability for two categories sharing the same cue distribution, such a pattern is evident in the production of Center of Gravity (COG) an acoustic measure used to describe sibilants (e.g., /s/ and /ʃ/). Here, the /ʃ/ category shows lower category dispersion within a talker while /s/ generally shows higher variability and greater category dispersion (Gunter et al., 2021; Newman et al., 2001).

Evidence for sensitivity to category dispersion is demonstrated in listeners' ability to categorize sibilant tokens. Newman et al. (2001) find that category dispersion within a talker predicted the speed at which listeners categorized, such that categories demonstrating wider dispersion (i.e., greater within-talker variability, /s/) resulted in slower and less consistent categorization compared to narrower categories (i.e., more consistent productions such as /ʃ/; see

also Clayards et al., 2008). However, exposure to high variability categories within a talker promotes greater generalization to novel tokens facilitating categorization of unheard items (Brosseau-Lapré et al., 2013; Zhao, 2010). Theodore and Monto (2019) demonstrated a similar pattern where listeners exposed to atypical VOT productions with either wide or narrow dispersion demonstrate different categorization functions with shallow functions in the former, but steeper functions in the latter. They argue that when listeners receive consistent input (low dispersion), the category boundaries are less fuzzy, and identification is more categorical. However, the distinctions between listeners exposed to narrow or wide distributions were attenuated over the course of the experiment as the cumulative distributional statistics between the two conditions converged, demonstrating flexibility in the cue-to-category mapping of a given talker over time.

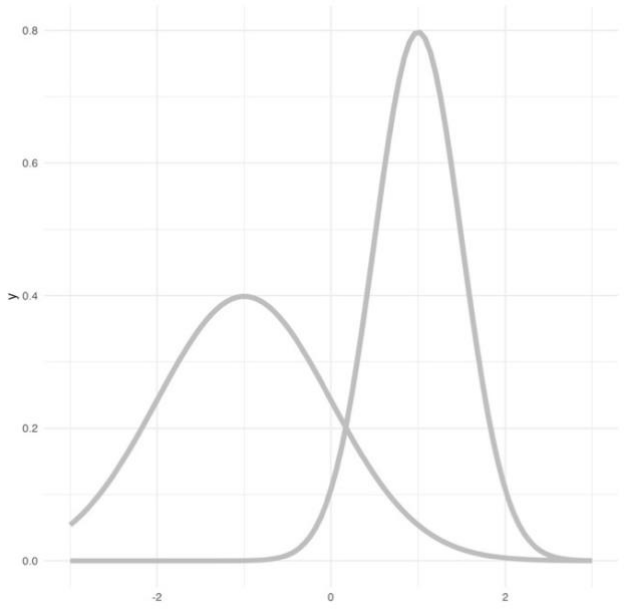

Figure 2.6 Example of distributional properties of two categories that vary in degree of dispersion. A realistic example might be extended to sibilants, where distributions of /s/ may look more like the wider dispersion category, and /ʃ/ may look more similar to the narrow dispersion category. Categories are distinguished by their mean and variance.

Such variability of an individual category (i.e., the dispersion or width of the category) may be driven by several factors including between-talker differences and within-talker

variability caused by phonological context, speech rate, social meaning, and so on. Distributional learning is thus pertinent to socio-indexical structure where listeners learn talker-specific and group-specific characteristics from distributions over the phonetic space (Foulkes & Hay, 2015; Kleinschmidt & Jaeger, 2015; Kleinschmidt, 2019). Indeed, there is evidence that listeners can learn the statistical distributions of linguistic features associated with group identities (e.g., Docherty, Langstrof, & Foulkes, 2013) and can track multiple talker identities in the input (Munson, 2011).

Additionally, listeners learn atypical productions for individual talkers from cumulative cues from the same talker over time (Theodore & Miller, 2010; Lai, 2021; Zhang et al., 2021). Lai (2021) further demonstrates a cumulative update mechanism where listeners integrate evidence from multiple talkers during exposure to atypical productions. Such a mechanism suggests that category dispersion as a function of individual talker identity may not be a stable property of a single talker but may be influenced by other similar talkers over the course of an experiment. Lai (2021) further illustrated that a cumulative updating mechanism may be modulated by socio-indexical cues to identity, where opposing indexical cues may block the integration of multiple talkers' cue distributions (see Section 6 for additional information). Overall, there is evidence suggesting that language users are attending to not only the overall distributions of a category and similarity of exemplars from experience but also the parameters of those distributions which are influenced by individual talker characteristics and the cumulative distributions of multiple talkers.

Thus, there is support for distributional learning of talker-specific characteristics, and initial evidence that such a mechanism extends to cross-talker variability and is modulated by socio-indexical factors. Distributional learning aligns with observations thus far, providing evidence that listeners learn group-specific characteristics through distributional input. Such a mechanism provides an account for the types of perceptual behavior in Section 4, including listeners' ability to track talker characteristics for evaluations of accentedness (Gunter et al., 2020) and dialect categorization (Clopper & Pisoni, 2004a-c; Clopper et al., 2006). Distributional learning governs listeners' ability to make inferences about the characteristics of a talker and their production system from the experienced distributions to guide speech processing.

5.2 Bayesian Decision-Making & Inferential Speech Perception

A key issue within distributional learning is that listeners must identify what phonetic variability is relevant to the linguistic message (e.g., distinguishes phones), what is relevant to the identity of the talker, and what variability may be incidental or noise. Several researchers have argued for Bayesian models of speech processing, where prior knowledge of a cue distribution's parameters (e.g., mean and variance) and underlying causes probabilistically guide speech perception and adaptation to variability (Clayards et al., 2008; Massaro, 1987; Massaro & Cohen, 1991, 1993; McMurray & Jongman, 2011; Jongman & McMurray, 2017; Kleinschmidt, 2016, 2019; Kleinschmidt & Jaeger, 2015; Liu & Jager, 2018; Norris & McQueen, 2008; Weatherholtz & Jaeger, 2016). While the exact implementations of models vary, the predominant framework posits that perceivers act as optimal observers (see also ideal adapters, recognizers, or perceivers), whereby listeners adopt the most optimal strategy for speech recognition and engage in speech perception as a process of inference under uncertainty (Clayards et al., 2008; Heald & Nusbaum, 2014; Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008; Nusbaum & Magnuson, 1997; Magnuson & Nusbaum, 2007; Pierrehumbert, 2016). Inference under uncertainty refers to a process "whereby listeners combine what they know about how speech is generated in order to recover (or infer) the most likely explanation for the speech sounds they hear" (Kleinschmidt, 2019:44), allowing listeners to resolve ambiguity in the speech signal. Under perceptual uncertainty listeners allocate credibility to different hypotheses and infer the most likely candidate based on prior experience and the most likely cause for the variability, thus recovering the intended message and the likelihood that such variability will be useful in future interactions with the talker (or other talkers).

As described above, some Bayesian models propose that listeners form a generative model of variability, which aims to capture the statistical structure of observed input, by tracking both the variable cue distributions and the causal links to the observed variability (Clark, 2013; Kruschke, 1996; Kleinschmidt, 2019). There are two fundamental problems to solve in a generative model: 1) inference of the most likely cause for an individual item (e.g., an experienced percept) and 2) the best model to capture the statistical structure of all experiences. The two problems cover significant aspects of speech processing—the first problem is related to

perception (i.e., identifying categories) and the second is related to adaptation (i.e., updating categories/perceptual recalibration). The generative model is informed by the prior probabilities of variability, drawn from previous experience with language use, which informs the likelihood of probable interpretations of the signal (e.g., the linguistic category) and their underlying causes (Kapatsinski, 2018; Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Kruschke, 2006; Liu & Jaeger, 2018). Generative models can be hierarchical and bidirectional (Clark, 2013) with listeners having higher-order top-down knowledge, including socio-indexical factors, that captures lower-order acoustic variability (Norris & McQueen, 2008; Kleinschmidt & Jaeger, 2015; Kleinschmidt, 2019; Liu & Jaeger, 2018; Weatherholtz & Jaeger, 2016). As such, listeners may infer causes relating to talker-specific patterns (Eisner & McQueen, 2005; Kraljic et al., 2008; Liu & Jaeger, 2018) or social context and group identity (Kleinschmidt, 2019; Weatherholtz & Jaeger, 2016). The parameters and nature of socio-indexical structure in inferential speech processing have been an underexamined question, however, such causal models would provide benefits for speech perception (Kleinschmidt, 2019; Liu & Jaeger, 2018), but also, crucially, allow language users to reproduce the sounds they hear for their social goals.

Kleinschmidt (2016, 2019) began to tackle the problem of the possible parameters of socio-indexical structure that are available to listeners in his computational-level theory of the ideal adapter model. In his work, he argues that listeners track individual talkers and social groups (e.g., dialect, gender, age) when they are informative for speech recognition, as evidenced by statistical conditioning of acoustic cues. Social grouping variables are highly informative and provide predictive power when the group demonstrates internal regularity (i.e., talkers systematically pattern together; Kleinschmidt, 2016). Informativity in his model refers to the degree to which a particular grouping factor is *relevant* to speech perception tasks, as modeled computationally (see Chapter 4 for more details). When socio-indexical factors are informative, listeners draw on these statistical contingencies to resolve ambiguity in speech, adapt to novel variation, and generalize patterns across talkers from a priori inferences about a talker based on experiences with similar talkers.

As an example, gender is informative of the acoustic cue distributions for /s/ and partitions the continuous variability into meaningful subsets, such that the mean and variance are reduced when we aggregate over each group (rather than the category as a whole). As such,

listeners will approach speech perception with a priori expectations about a given talker's production of /s/ based on inferred gender identity—a more fronted /s/ (higher mean COG) for a woman and a retracted /s/ (lower mean COG) for men. If a talker deviates substantially from their inferred social group, listeners will adapt and learn talker-specific patterns. Thus, for Kleinschmidt (2019) socio-indexical structure refers to the statistical conditioning of raw cues by social groups which act as the lower bound from which talker-specific learning occurs. The ideal adapter model describes listeners' behavior as optimally driven, and as such listeners should draw on the causal relationships of social information and phonetic variability a priori to make online speech processing more efficient. To move towards a comprehensive theory of socio-indexical structure in inferential speech processing models, researchers need a more comprehensive descriptive characterization of the informational assumptions about what listeners track, the levels of relevant socio-indexical factors and group characterizations, and variable prior experiences.

An initial assumption about socio-indexical structure is the nature of the social dimensions listeners track depends on listeners inferring how talkers should be grouped (Kleinschmidt, 2019). Currently, Kleinschmidt's (2019) model bifurcates such inferred social dimensions into group-informative and talker-informative components, where the group-informative variable relies on the group demonstrating internal regularity across talkers. Such a model, however, does not specify how differences of between and within-talker variability contribute to shaping the distributional parameters and aggregation that inform listeners' beliefs. However, the ideal adapter model predictions of socio-indexical structure hinge on such a dichotomy at various stages of speech processing.

First, listeners draw on the causal links to infer the intended linguistic category, with group identity aiding in predicting and parsing the signal. When a contrast's variability is conditioned by group, prior experience with other talkers will be more relevant for resolving perceptual ambiguity. If it is not conditioned on group, and individual talker identity is highly predictive, prior experience with other talkers is less informative in the initial resolution of ambiguity. When a talker has novel production patterns, listeners must infer whether such variability is caused by characteristics of the talker. In highly talker-informative categories, adaptation occurs. It's less clear from the model whether listeners make the same inference,

absent of other disambiguating information, that the variability is likely characteristic of the talker when the category is conditioned on group.

Ultimately, a critical distinction for cross-talker variation stems from generalization behavior, which hinges on group-informative and talker-informative separation. Ideal adapter models posit that listeners learn in talker-specific ways when the long-term experience with cue distributions demonstrates little statistical relationship with grouping factors, and on the other hand update category representations more globally when cue distributions are more strongly conditioned on group-level factors (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). The social dimensions outlined here are undoubtedly important, however, the make-up of social groups and the relationship between individual talkers and groups requires further refinement and interaction with sociolinguistics; I will return to this point in more detail in Section 7.

A second assumption warranting further characterization is how prior experience shapes listeners' inferences. Currently, prior experience is treated in a one-size-fits-all manner to gain insight into structured variation. However, at the level of individuals we expect complex social experiences will be reflected in different experiences with and representations of talker variability. For any given theoretical account, the 'experience' may be underspecified and vague statements or a statistical description based on all of American English for a given category or categories. In other words, researchers make a simplifying assumption that all experience with variability will inherently encompass a broad perspective of American English, and it is from this broad representation, listeners draw inferences, and not a more limited (e.g., single region) or imbalanced experience. Kleinschmidt (2019: 62) further acknowledges that "every listener's experience with talker variability will be different, and so a variable that is informative in one listener's experience may be irrelevant in another's". In other words, listeners learn different generative models of variability in the world based on their communities, their mobility, and other individual differences. However, listeners' responses to dialectally conditioned variability suggest a degree of shared knowledge of variability despite different experiences with imperfect and noisy distributions (as illustrated in Section 4 above). Ideal adapter models aim to characterize the input and output of listener behavior, and thus more description of the nature of

variable experience (i.e., the input) is needed to understand listener behaviors (i.e., the output) more comprehensively; I will return to this point in more detail in Chapter 4 and 5.

Finally, a critical component of socio-indexical structure in speech inference is the nature of what listeners track in relation to socio-indexical factors. The formalization of socio-indexical structure currently holds that a given talker's accent can be operationalized as an aggregate of the raw distribution of cues for a single category (or group of categories). By extension, a social group can be characterized as an aggregate of the raw distribution of cues for a given category (or group of categories). Kleinschmidt argues that socio-indexical structure in ideal observer models is different from other forms of category structure such as correlations of a cue across categories (e.g., talker-specific mean VOT for /b/ and /d/), and relationship among cues within a single category (VOT and F0 for voicing; Clayards et al., 2008; Idemaru & Holt, 2011, 2014). Such a holistic perspective of categories ignores the nuances of the type and internal organization of variability (within a category, talker, or group) exemplified in sociophonetic studies. As illustrated above, for example, vowels are systematically organized, and individual vowel categories are not entirely independent from one another. Such organization may be demonstrated in several ways, including the dependency between categories (Tamminga et al., 2016), and the linear relationship between a single cue and a phonetic dimension (Chodroff, 2017; Chodroff & Wilson, 2020, 2022). Such internal regularity between categories has additional ramifications for how listeners recover intended categories and resolve ambiguity across talkers. While Kleinschmidt (2019) acknowledges that socio-indexical structure is complementary to other specific components of category structure, the model lacks a clear explication of how they complement one another or why such category structure is distinct from socio-indexical structure and at what stage more specific learning occurs. The assumptions about the nature of what listeners track related to socio-indexical factors warrant further interaction with (socio-)phonetics and integration into descriptions of how listeners infer group and talker structure; I will revisit this point in Section 7 below and in again in Chapter 5.

From the above, it is clear that several open questions remain about what listeners track, under what contexts, and when such information is generalized to other talkers. Together such complexities illustrate the tension which Bayesian models seek to address: specificity on the one hand (e.g., experience, phonetic dependencies, etc.) and generalization on the other (e.g., across

talkers, contexts). Theories based on Bayesian models broadly speaking elucidate much of the stability and flexibility of the system by outlining how listeners leverage prior experience to make probabilistic inferences from input. However, a major gap in Bayesian models is the thread linking specificity to socio-indexical structure—that is, how much specificity is tracked and learned at different levels of socio-indexical structure (e.g., talkers, dialects, etc.)? The domain of perceptual learning speaks to some of these questions, highlighting variable listener behaviors in the face of different types of variable input. In the section below, I will review the current literature on perceptual learning focusing on the ways in which linguistic knowledge guides talker-specific learning and cross-talker generalization.

6 Perceptual Learning

As previously described, perceptual learning can be explained via processes of distributional learning and Bayesian inference (Clayards et al., 2008; Kleinschmidt & Jaeger, 2015; McMurray & Jongman, 2011; Norris & McQueen, 2008). Perceptual learning variably refers to the process whereby listeners learn novel phonetic categories (Logan, Lively, & Pisoni, 1991), recalibrate cues to existing categories (Clayards et al., 2008; Idemaru & Holt, 2011), adapt (broadly) to novel talkers and accents (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004), and adapt to talker segmental variation (Cutler et al., 2008, Norris et al., 2003; Samuel & Kraljic, 2009). The primary focus of this dissertation is concerned with how listeners adapt to cross-talker segmental variation, and thus this section will emphasize this subsection of the literature on perceptual learning and will refer to it simply as perceptual learning.

Perceptual learning studies regularly make use of lexically guided perceptual learning paradigms (Norris et al., 2003) to assess learning of talker-specific phonetic distributions. In these paradigms, listeners are exposed to a novel production that is (typically) phonetically ambiguous, embedded in disambiguating lexical contexts. The lexical context allows listeners to scaffold learning by recovering the intended category of the ambiguous production which subsequently aids in the remapping of linguistic categories. Learning is demonstrated when listeners shift their phonemic boundaries (Cutler et al., 2008; Kraljic & Samuel, 2005; Norris et al., 2003) or show increased word endorsement rates after exposure (Maye et al., 2008;

Weatherholtz, 2015). Some work has demonstrated that listeners show phonetic retuning without such top-down influence using only nonce words or syllables, supporting an underlying distributional learning mechanism (Chládková et al., 2017; Munson, 2011), but such work is still relatively limited.

Much of the perceptual learning research has examined constraints on learning and the conditions under which listeners generalize novel phonetic patterns to other related phonetic categories and talkers. Overall, such work highlights the general debate between specificity on the one hand and generalization on the other, observing both a high degree of specificity in learning individual talkers (Kraljic & Samuel, 2006, 2007; Norris et al., 2003; Nygaard & Pisoni, 1998; Samuel & Kraljic, 2009; Theodore & Miller, 2008) and types of variability (Babel et al., 2021; Dahan et al., 2008; Idemaru & Vaughn, 2020) alongside generalizing such patterns across talkers and contrasts sharing phonetic dimensions (Kraljic & Samuel, 2006). In the next section, I will review the constraints on learning, focusing on addressing the question of what listeners track, followed by a review of the constraints on when cross-talker generalization occurs.

6.1 Constraints on Learning

A large body of work in perceptual learning supports Bayesian inference, whereby listeners recalibrate phonetic categories to talker(s) when the short-term cue distribution deviates from their long-term representations for the category. Such recalibration is sensitive to previous experience, which may produce asymmetries in listeners' perceptual learning behavior across and within contrasts. For example, recent work by Babel et al. (2021) demonstrates that listeners are sensitive to typological regularities in voicing and devoicing. Here, they train listeners on either a typologically common change, devoicing of /z/ to [s], or a typologically uncommon change, voicing of /s/ to [z]. In line with this typological pattern, they find that listeners don't learn the typologically irregular pattern from exposure, and instead demonstrate a more global relaxation of criteria for /s/ tokens, accepting pronunciations outside of both the exposure pattern and typical /s/ productions. On the other hand, listeners show targeted learning for the typologically common devoicing pattern of /z/, accepting more devoiced /z/ tokens but not pronunciations they didn't experience during exposure. Babel et al. (2021:50) argues this is the result of prior knowledge about the commonality of devoicing in English, and that "subphonemic

changes with which listeners have linguistic experience facilitates targeted adjustments, while novel changes seem to spur more global criteria adjustment". Other work suggests that listeners are more rigid in targeted adjustments to categories when exposed to novel shifts outside category typicality (Babel et al., 2019; Kleinschmidt & Jaeger, 2015; Sumner, 2011) and have difficulty learning joint cue distributions that are positively correlated in the environment but negatively correlated in the experiment (Idemaru & Holt, 2011). Such evidence points to experience with common patterns and ranges of variability along specific cue dimensions from which listeners form expectations about the boundaries and likelihood of variable forms.

In support of Bayesian inference, there is evidence that listeners attend to the underlying cause of variation during perceptual learning tasks. Some work demonstrates that listeners are unlikely to learn talker-specific details when they are provided evidence that the novel production is caused by incidental factors. Such incidental factors may be inferred by the presentation order of the stimuli where 'typical' clear productions are followed by atypical productions (Sumner, 2011; Kraljic et al., 2008; Kraljic & Samuel, 2011) or through explicit visual disambiguating information (Kraljic et al., 2008; Liu & Jaeger, 2018). Kraljic et al. (2008) demonstrate that listeners do not learn an ambiguous production of /s/ (between /s/ and /ʃ/) when the talker is holding a pen in their mouth. They argue that in such cases the listener is likely to infer that the percept is incidental and not a characteristic feature of that talker's speech, and thus discard and ignore the information from exposure.

In a follow-up, Liu and Jaeger (2018) argue that listeners are able to maintain several inferences of potential explanations for variability when it is causally ambiguous rather than discarding the experience. A percept is considered to be causally ambiguous when there is no discernable evidence of the underlying cause of the percept, if it is later disambiguated it becomes causally unambiguous. Listeners maintain information if the cause of the percept is ambiguous and they are uncertain about whether to attribute it to incidental causes (e.g., pen in the mouth) or characteristic causes (e.g., talker-specific characteristics). In the case of the pen in the mouth, the initial percept is causally ambiguous because the atypical percept (e.g., dino[s]aur → dino[ʃ]aur) could be attributed to either the pen in the mouth or talker-specific characteristics (i.e., idiosyncratic causes). If the listener is then exposed to typical productions with the pen in the speaker's hand, it is no longer ambiguous, as the change in percept and change in pen

position indicates a causal relationship. If the atypical percept continues despite the pen in the hand, it disambiguates the pattern and suggests it is characteristic of the talker (Liu & Jaeger, 2018). Such work suggests that listeners overall retain information for uncertain input and may integrate such cues after disambiguation occurs. This is further supported by recent work that the order of the stimulus does necessarily block learning, and instead, listeners appear to be sensitive to cumulative cues in incremental updating (Cummings & Theodore, 2023; Tzeng et al., 2021; Lai, 2021). The role of causal inferences for socio-indexical information, such as dialect, during perceptual learning has remained unexplored (though, see Chapter 6). However, some work suggests that if listeners are faced with variability attributable to dialectal factors, or categories that are broadly variable, listeners may be less likely to show learning (Kraljic et al., 2008; Kataoka & Koo, 2017).

In a similar vein, Kraljic et al. (2008) argue that learning only occurs when the system has no viable alternative solution for experienced variation and learning is driven by idiosyncratic variable productions that are not contextually dependent (i.e., occur across the category). As support for their argument, Kraljic et al. (2008) provide evidence that listeners do not learn an ambiguous /s/ percept (i.e., shift from /s/ towards /ʃ/) before [tr] clusters, a common pattern across regional dialects. They argue that such contextually dependent variation can be explained by listeners by assigning the ambiguous percept to features in the phonetic context rather than characteristic of the talker's /s/ category more broadly, despite the link to socio-indexical causes. While not embedded in a Bayesian framework, such an account is echoed in some models of Bayesian inference of socio-indexical structure, where listeners' inferences are drawn from a raw cue distribution rather than contextually specified variants (Kleinschmidt & Jaeger, 2015; Kleinschmidt, 2019; Weatherholtz & Jaeger, 2016) and thus talker-specific and dialect patterns are represented as an aggregate over all contextual variation.

Yet, there is evidence that listeners track separate cue distribution statistics for instances of a single category in different contexts for a given talker (Dahan et al., 2008; Idemaru & Vaughn, 2020). For example, Dahan et al. (2008) demonstrate that listeners learn about context specific raising of /æ/ before /g/ (e.g., bag) to guide speech processing of both raised and non-raised /æ/ contexts. Integration of contextual variability is described in other theories of representation and perception more broadly (Apfelbaum et al., 2014; Cole et al., 2010; Jongman

et al., 2000; McMurray & Jongman, 2011; Pierrehumbert, 2016) with an argument for similar mechanisms where representations are built from experience[3]. As an extension of Bayesian models, one may hypothesize it aids in listeners' acquisition, as they may infer from categories that have a bimodal distribution from such contextual variation, are likely to be the result of two separate targets (i.e., allophones; see Kapatsinski, 2018) aiding in learning the phonological system. Such a mechanism may be applied to socio-indexical structure as well, where listeners learn such contextually specified variation. Thus, the blocking of perceptual learning observed by Kraljic et al. (2008) may be explained by listeners' existing experience with the variant through talkers from regional dialects with the pattern (e.g., Baker et al., 2011) rather than provide overwhelming evidence that listeners will resist updating representations in such contexts.

Rather than contextually bound blocking of variants, it's possible the short-term distributional characteristics did not deviate from their long-term experience enough to warrant recalibration of the category. There is potential evidence for this fact, as Kraljic et al. (2008) demonstrate listeners can replicate the pattern in production when asked to imitate the voice, regardless of whether they have the variant in their dialect, suggesting the variability is not discarded altogether. Thus, it's likely that listeners are sensitive to *within*-talker contextual variability and expectations, and retuning may be relative to expectations of category structure alongside expectations of cross-talker variability. It's unclear how such components can be incorporated into an account of socio-indexical structure and Bayesian inference, and while such fine-grained phonetic conditioning is outside of the scope of this dissertation, such contingencies illustrate the complexities of how much specificity listeners track in explaining category variance from socio-indexical factors (see Apfelbaum et al., 2014 for a similar discussion).

Research demonstrates listeners are likely to generalize to other phonetic environments sharing the same phonetic dimension (e.g., voicing), providing evidence that listeners learn talkers' cue distributions across categories that are not entirely independent from one another.

[3] There is a long research tradition that examines coarticulation in perception that takes a different theoretical perspective of gestural (e.g., Fowler, 1994) and feature parsing (e.g., Gow, 2003). Such theoretical perspectives are outside of the scope of this dissertation, as such I focus on cases integrating distributional learning or exemplar theoretic perspectives into context-specific variation. Covariation accounts may similarly be identified as a form of feature parsing accounts, but this is not essential for such information to be integrated into beliefs.

Such contingencies are beneficial to talker-specific learning, where such regularity within cue distributions aid in making learning tractable for a given talker (Chodroff & Wilson, 2020). For example, Kraljic and Samuel (2006) illustrates that when listeners are trained on ambiguous percepts of /t/-/d/ that bias listeners to resolve /t/ or /d/, listeners will generalize to a category that shares the voicing contrast with the exposure category (e.g., /d/ → /b/, /t/→ /p/). However, such generalization may be constrained to categories where the acoustic cues to identity are shared within the segment, and not distributed over neighboring sounds (e.g., voicing vs. place of articulation; Idemaru & Holt, 2014; Mitterer & Reinisch, 2017).

Chládková et al. (2017) demonstrate that Greek listeners adjust their boundary between /i/ and /eɪ/ in the direction of exposure (/i/ lowering or /eɪ/ raising) and generalized the pattern to the talker's complementary /u/-/o/ boundary. Such findings suggest that listeners may dynamically adjust boundaries across contrasts with shared contrasting dimensions (i.e., phonological features). Learning talker-specific characteristics is thus likely to generalize to unexperienced categories that share contrastive phonetic dimensions. Such knowledge may be elucidated by ideal adapter principles whereby listeners generalize to perceptually or acoustically similar ranges (Reinisch & Holt, 2014) or may otherwise reflect some tracking of relationships across categories for a given talker. Thus, cross-category relationships within and between talkers highlight an additional dimension along which listeners track talker-specific (Chodroff & Wilson, 2020) or, potentially, group-specific variability (see Chapter 5 for further discussion).

Work in perceptual learning has predominately examined consonantal variation, whereby listeners are given ambiguous productions along a single dimension (e.g., Center of Gravity for sibilants, or VOT for stops), thus, it is an open question whether such rapid adaptation occurs for vowels to the same degree and under similar constraints. Vowels have significant differences from consonants that make them an interesting test case. In particular, vowels carry a high degree of talker-specific information given their spectral quality compared to, for example, stops. On the other hand, they also carry a large degree of indexical information about talkers' social backgrounds and group-level attributes. As such, vowels are both talker-specific and group-specific. Primarily, work examining vowel variability in lexically guided perceptual learning has examined cross-category remapping (e.g., /æ/ is remapped to /ɛ/) in novel chain shifts, whereby multiple vowels are affected in the shift (Babel et al., 2019; Maye et al., 2008; Weatherholtz,

2015). This methodology can thus be distinguished from the predominate work in consonantal variation that examines the restructuring of phonetic boundaries after exposure to ambiguous items and represents a more general learning of a talker's 'accent' across all or part of the vowel space which may occur even when the voice is not perceived as socially favorable (Babel et al., 2019). Results have demonstrated variable findings with arguments that adaptation is context and vowel-dependent (Maye et al., 2008) or the result of a general broadening of perceptual space accepting a wider range of variability (Weatherholtz, 2015).

In particular, Weatherholtz (2015) demonstrated that when exposed to a novel back vowel lowered shift, listeners generalized to a novel back vowel raising shift they had not previously been exposed to. However, this is not true in the other direction, such that when listeners were exposed to back vowel raising, they did not generalize to back vowel lowered variants (Weatherholtz, 2015). On the other hand, Maye et al. (2008) demonstrated direction-specific learning when listeners were exposed to front vowel lowered shift. Such findings generally demonstrate that listeners attend to the direction of shifts and learning may be constrained therein. While the role of prior experience for vowel shifts has been unexplored, it's plausible that constraints may be driven by listeners' prior experience with typological regularities of directions of shifts (e.g., raising/lowering). Analogously to sibilant learning of typological voicing patterns (Babel et al., 2021), we might find that the lowering of front vowels is within the range of experience for listeners, resulting in targeted adjustments as in Maye et al. (2008) but back vowel lowering may result in more global adjustments as in Weatherholtz (2015); I will return to this point again in Chapter 5.

Additionally, there is some evidence that listeners draw on prior experience with variability and demonstrate asymmetrical learning across categories distinguished by within-category variability. Kataoka and Koo (2017) demonstrate that individual vowel categories demonstrate different malleability in adaptation, such that high variability vowels (e.g., /u/), show less evidence of learning than low variability vowels (e.g., /i/) in contextually bound shifts (i.e., before liquids). This finding is generally counter to the predictions that some ideal adapter models make (e.g., Kleinschmidt, 2019) since vowel categories are broadly hypothesized to be adapted to robustly due to the degree of cross-talker variability. As such, the prediction in such models is that listeners are likely to demonstrate more complete adaptation when a contrast is

likely to vary across individual talkers. Thus, we would predict that learning should occur equally across the two categories, as prior experience with a talker is expected to be less informative.

However, the results could still be explained by prior experience and a more nuanced understanding of asymmetries among vowel categories. In particular, American English listeners are likely to have extensive experience with /u/-fronting across dialects providing again no reason to adapt to the short-term experience. In such a case, the fact that /u/ was fronting in a context they had, perhaps, not experienced may have been less relevant to the overall tendency of /u/ to front. Alternatively, listeners may see more reason to update narrow categories (as in the case of /i/) but may more readily associate tokens with /u/ due to the broader range of variability, in line with the observations of Cohen et al. (2001), resulting in a broader acceptance of /u/ but overall limited retuning. An additional complication of Kataoka and Koo (2017) is that variability is broadly construed, encompassing multiple potential causes, including but not limited to, social factors and phonetic context. Overall, it's unclear whether different causes of variability contribute equally to perceptual flexibility, and how socio-indexical factors may contribute to such flexibility. Future work should ascertain why such adaption asymmetry occurs, disentangling different sources and types of within and between-talker variability (see Section 7 and Chapter 6 for additional discussion).

6.2 Constraints on Generalization

Scholars have argued that perceptual learning is a low-level process and talker-specific (Eisner & McQueen, 2005). More recently, work has begun to address under what conditions learned patterns generalize to novel talkers with findings suggesting acoustic similarity (e.g., Reinisch & Holt, 2014; Xie & Meyers, 2017), contrast type (Kraljic & Samuel, 2006; Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015), learned socio-indexical causes (Kleinschmidt, 2019; Weatherholtz & Jaeger, 2016), and variability in exposure (Sumner, 2011) predict generalization patterns. Furthermore, some work posits that talker-specific learning occurs in cases where there is a greater degree of spectral information contained in the contrast (e.g., sibilants) than for other categories (e.g., stops, Kraljic & Samuel, 2006). However, current work thus far demonstrates evidence against this hypothesis showing that exposure to novel

vowel shifts generalize across talkers robustly compared to fricatives and stops, which demonstrate variable generalization patterns (Eisner et al., 2013; Kleinschmidt & Jaeger 2015; Kraljic & Samuel, 2006, 2007; Tamminga et al., 2016; Van der Zande et al., 2014).

Patterns of generalization that emphasize talker acoustic similarity have often operationalized such similarity through the use of same or different gender pairing to the exposure talker (Weatherholtz, 2015; Reinisch & Holt, 2014; Witteman et al., 2013) or through the acoustic similarity within the segment itself (Lai, 2021; Tamminga et al., 2016). In the former case, such operationalizing of acoustic similarity may inadvertently collapse acoustic similarity and social identity (Tamminga et al., 2016; Lai, 2021). As such, the results of variable generalization patterns may be explained by listeners' knowledge of socio-indexical structure rather than a low-level acoustic similarity. For example, Kraljic and Samuel (2007) demonstrate that generalization occurs across gender pairs for voicing shifts among stops but not for sibilants, where learning is talker-specific or constrained to same-gender pairs. Weatherholtz (2015) likewise demonstrated generalization across different-gendered pairs following exposure to a novel vowel shift across different gendered pairs. In terms of socio-indexical structure, listeners' prior experience may promote restrictions to sibilant generalization to members of the same gender group since sibilant productions demonstrate robust gendered patterns. Recent work by Lai (2021), however, demonstrates robust generalization across genders for both stops and sibilants which remains stable across different manipulations of the acoustic properties of the target sound. However, manipulating visual and vocal cues to the social identity of the generalization speaker results in attenuated within-gender pair generalization suggesting that socio-indexical structure may provide gradient constraints on the magnitude of generalization rather than overall blocking (Lai, 2021).

Such socio-indexical modulation in perceptual learning shows that listeners are sensitive to the distributional make-up of social groups which inform their perceptual learning behavior and may be seen in gradient rather than categorical shifts in listener behavior. Such evidence is still limited and contends with several other studies that have demonstrated constraints to such generalization and may interact with other facets including perceptual similarity. Understanding of how such perceptual behavior aligns with talker variability in the real world is limited, making it challenging to distinguish between socio-indexicality or other perceptual mechanisms driving

generalization. Additionally, despite vowels being widely recognized as paramount examples of socio-indexical variation, there is a dearth of literature examining the generalization of vowels. Looking at categories beyond stops and sibilants may further elucidate the role of socio-indexical structure in perceptual learning.

Overall, the perceptual learning literature offers a complex picture in terms of the specificity and nature of variability at an individual talker level. Some perceptual learning suggests that listeners are only likely to retune when the input is a holistic category shift and when there are no alternate causes to explain the variation (Kraljic et al., 2008; Liu & Jaeger, 2018), resulting in talker-specific learning. However, there is increasing evidence that listeners learn contextual (Dahan et al., 2008; Idemaru & Vaughn, 2020) and cue-specific (Harmon et al., 2019; Idemaru & Holt, 2014; Reinisch et al., 2014) patterns for an individual talker. Furthermore, evidence of constraints for the specificity of learning may be attributable to listeners' long-term representations of learned variability (e.g., /str/ and /u/ fronting), which warrants examining production patterns to understand the variability with which listeners have experience to inform perceptual learning (see also Babel et al., 2021). Additionally, the why and how of listener generalization to other talkers is still largely unclear, and at what level of socio-indexical structure guides this process. Turning to sociolinguistics for a more nuanced understanding of socially meaningful variation may further aid in refining models of socio-indexical structure in Bayesian inference and the predictions therein.

7 Socio-Indexical Structure in Production

A necessary first step to building a model of socio-indexical structure of variability is to understand the sources of variability in the signal. Currently, socio-indexical structure in Bayesian models thus far identifies social groups and individual talker identity as sources of variation, however, the nature of such groups is relatively vague and encompasses more than one type (e.g., between and within) and source of variation. Not all variation is socially meaningful, and some differences across talkers may be the result of physiological differences or shared linguistic sources of variation that may be, theoretically, shared by a speech community (e.g., coarticulation; see also Ladefoged & Broadbent, 1957 for similar taxonomy). Much of the discussion thus far has not disentangled these sources of variation from the social. While not

mutually exclusive, the physiological, social, and linguistic sources may result in different distributional properties. Given the interest in the social, I will not review the wide range of physiological factors contributing to variability and instead will focus on the relationship of individuals to their social groups which will also encompass a fair degree of internal linguistic sources as well.

There are two assumptions that require a deeper interaction with sociophonetics to fully integrate socio-indexical structure into Bayesian models that stem from the dichotomy of talker-specific and group-specific patterns. The first assumption is that variability is conditioned on social groups only in so much as they are internally homogenous and provide relevant information for speech processing, either to parse the signal or to identify social characteristics (Kleinschmidt, 2019; Kleinschmidt et al., 2018). Namely, when individuals are meaningfully organized into social groups, there is little between-talker variability within social groups which provides statistical regularity to the acoustic distributions for listeners to learn and leverage during speech processing. The second assumption is perhaps a simplification of computational models (e.g., Kleinschmidt, 2019), where 'talker-specific' information collapses between-talker and within-talker variability. Theoretically, when cross-talker variability is high listeners can resolve ambiguity induced by such variation through learning an individual talker or through their perceived group identity. However, there is little discussion about where and how within-talker variability fits into this framework. This simplification may have been requisite for building the theoretical description but is of critical importance for furthering our theories about how listeners make sense of talker-specific or dialectally conditioned categories in speech processing. These two assumptions are inherently linked, as assumptions of group homogeneity may be seen as counter to some types of individual talker variability (i.e., style), while other forms of individual talker variability may largely be reflected across talkers within the group (e.g., phonological variation). If listeners are tracking and exploiting these statistical relationships for speech processing, then it's critical we have a more thorough understanding of each component.

While this dissertation will not address the full scope of these issues, in the following sections I provide an overview of the nuances of individuals and their social groups, again focusing primarily on dialect groups, and outlining how different patterns of intra-group (i.e.,

between talkers in a group) and within-talker variability may ultimately shape distributional properties which act as the input for learning. Given the various sources and outcomes of variability, I provide a taxonomy of variability that outlines what I believe to be the core components in need of further characterization. Following this overview, I suggest how our current models of socio-indexical structure can be informed by current discussions in sociolinguistics as to how listeners learn socio-indexical structure and drive inferential processes of speech perception.

7.1 Nature of the Group: Speech Communities & Heterogeneity

Within much of the literature reviewed thus far, I have referenced 'social group' and 'dialect' in relatively vague ways. Indeed, much work within phonetics and psycholinguistics have (rightfully) grouped talkers by practical terms with various macro social categories, including broad geographic regions (i.e., dialects), and as such the terms may be minimally defined. Of course, geographic region and other macro social categories have been established in sociolinguistics to correlate strongly with phonetic variation (e.g., Labov et al., 2006; Labov, 2001), making such choices valid for many research questions. In practice, the practical identification of external social groups as a unit for quantitative sociolinguistic analysis is generally linked to a division in geographical space (Horvath & Horvath, 2003; Gumperz, 2009). However, typically some action is taken from the outset to ensure that the talkers encompass the range of possible variations within the community. Further, a great deal of theoretical work in sociolinguistics has sought to define social groups that are meaningful organizational units of society and has wrestled with the representativity and limitations of macro social categories and speech communities (Bucholtz & Hall, 2005; Eckert 1988, 2000, 2012; Eckert & McConnell-Ginet, 2012; Milroy & Milroy, 1985, 1992; Rickford, 1986). Overall, this work has pointed to the complexities of not only identifying and delimiting meaningful groups according to geography but also the complexities of their internal organization, including subgroups and individual talkers.

In the chapters to follow, I will largely be following the same practical categorization of talkers into dialect areas as a starting point for the central questions outlined in Chapter 1. However, in this section, I aim to detail some of the potential concerns and issues that theories of

inferential speech processing must reconcile when discussing geographical dialects in theories of socio-indexical structure. Given the interest in dialect areas as a social group, I will primarily focus on geographic space and its connection to *speech communities* in the section to follow. In addition, I address one primary issue of internal group organization throughout this dissertation, despite the practical identification of dialect areas: the role of individuals and the extent to which they reflect the patterns of their dialect areas. As such, following the background on social groups, I will describe the role of individuals in more detail.

Insights into social groups can be gained by examining the theoretical and practical identification of communities and dialects in sociolinguistics. Much of this work has emphasized the critical role of the *speech community* and geographical space, which, crucially, can be defined and described in several ways. The speech community can refer to an aggregated group of people who share communication and norms between members (Labov, 1972) and/or the frequency of interaction (Gumperz, 1968). One of the core principles of quantitative approaches to variation has been to examine the speech community as the primary social group of interest under the guiding belief that the speech community is not merely a correlate of the linguistic system, but a core component of language use. Consequently, it has been suggested that the community, and not the individuals' grammar, be the object of linguistic study (Meyerhoff & Walker, 2013; Labov, 1972). Labov (1972: 120) further argues that norms may be observed in "the uniformity of abstract patterns of variation which are invariant to particular levels of usage". The emphasis on the aggregate group be described as the 'homogeneity assumption' (Wolfram & Beckett, 2000), which encompassed the belief that individual talkers' are not meaningful units of analysis in describing sociolinguistic variation so long as the social groups are meaningfully different. Such a definition may initially align with the identification of structured variation within a 'social group' or 'dialect' in phonetics and psycholinguistics. However, the speech community is not without variation in linguistic systems and may vary across several dimensions.

Speech communities are internally diverse in several ways, as talkers are differentiated by other social factors including gender, age, status, and social standing (Labov, 1972, 1994, 2001 among others). The speech community is often operationalized by a division in geographical space for practical reasons (Horvath & Horvath, 2003; Gumperz, 2009) and is often

a unit for placing the study itself rather than placing the talkers (Eckert, 2000). Thus, it's widely recognized that broader geographic dialect areas will be heterogeneous, but in systematic and quantifiable ways. Weinreich, Labov, and Herzog (1968) describe this as 'orderly heterogeneity', an observation that linguistic variation is systematically conditioned by intra-community social stratification and internal linguistic conditions (i.e., 'structured variation'; see also Labov, 1966, 1972; Wolfram, 1969). The geographic community can vary by size, from larger regional locations common in dialectology (e.g., Southern U.S., as in Labov et al., 2006) or in terms of large urban areas (e.g., New York City as in Labov, 1966 and Detroit as in Wolfram,1969). Geographic space is also reified as a social category across researchers in dialectology and sociophonetics as central to linguistic variation (Britain, 2013; Horvath & Horvath, 2002, 2003; Labov, 1980; Labov et al., 2006). It is thus critical that our theories of socio-indexical structure in speech processing explicate expectations of heterogeneity, both in how we identify socially structured variability in production, and how listeners make sense of it. Currently, this is a large gap in the theoretical assumptions of inferential speech processing, where dialect areas are a single undifferentiated aggregate with limited hierarchical or sub-group organization and reflect limited between-talker variation within a given group.

We can consider how dialect areas may reflect dependencies between subgroups at different scales, including the relationship between the individual and smaller communities. Horvath and Horvath (2003:144) argue for research examining the geolinguistic scale of variation, which ultimately captures the "nested hierarchical relationships as one move from the individual member of a speech locality to the speech locality, to the region, to the nation, and finally to the supranational scale". Through examining the nature of nested hierarchical relationships, they argue, linguists can understand the degree of universals in language more broadly as well as a more comprehensive understanding of variability. Yet, much of the work in sociophonetics has examined only individual layers of this problem, either focusing on the local (e.g., a city), the region (e.g., North/South), or the individual, rather than the nested or global nature of different varieties and linguistic variation.

While Horvath and Horvath (2003) were generally referring to geolinguistic scale for understanding variability, universality, and language change, the theoretical notion of nested geographical models of social and linguistic patterns has implications for how we conceptualize

socio-indexical structure, and the knowledge language users have about these nested relationships of talkers. Exemplar models adequately account for these nested relationships, by allowing for complex associations between social and acoustic information. However, processes of Bayesian inference that rely on social groups (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016), generally treat social groups as homogenous and make no reference to their nested or complex relationships. Yet, these groups are fundamentally nested products of geographic communities, that warrant further investigation. While there is some indication that socio-indexical structure is informative in the interaction between gender and dialect (Kleinschmidt, 2019), the hypotheses about how listeners draw on this nested structure for inferences remain unclear. This dissertation does not directly investigate all layers of nested divisions of geographic space, but it will begin to probe the relationship between individuals and their broader dialect areas. Thus, this dissertation speaks to the extent to which dialect groups diverge from the higher-level of national geographic location (U.S.) and the extent to which individuals fit within the regional geographic scale (e.g., North/South/West). In the following section, I will outline some of the current outstanding issues about how individuals are thought to reflect their dialect areas.

7.2 Nature of the Individual: Individuals Within Groups

A central issue in sociolinguistics for some time has been identifying the relationship between social groups and individuals (Bayley & Langman, 2004; Benor, 2008; Carpenter & Hillard, 2005; Forrest, 2015; Guy, 1980; Gumperz, 2009; Horvath & Horvath, 2003; Mendoza-Denton et al., 2010; Meyerhoff & Walker, 2007; Milroy & Margrain, 1980; Milroy & Milroy, 1985; Tamminga & Wade, 2022). And, indeed, this has been an underlying debate in linguistic theory more broadly (see e.g., Weinreich et al., 1968 for discussion). Such work has emphasized, on the one hand, the uniformity of abstract patterns of variation (e.g., vowel systems, Labov, 1972), and on the other, the agentive and dynamic use of phonetic variants by individuals in constructing identity (e.g., Eckert 2008, 2012). Chodroff and Wilson (2022) suggest such a division may comprise two ends of a continuum where on one end individuals are a maximal reflection of their speech community's behavior, and on the other hand, are fully agentive and not inherently bound by the patterns of their communities. The literature reviewed here will

follow this division, examining either the dependencies of variable (phonetic) forms or tracing the range of phonetic variability within talkers across contexts. Building on the ideas of socio-indexical structure outlined thus far, this section illustrates the tension between social groups and individual talkers drawing on concepts in sociolinguistics and phonetics.

Such debates are relevant to current theories of speech processing and highlight the dynamic and complex ways individuals pattern according to their respective groups. These two facets highlight two major assumptions that Bayesian models of socio-indexical structure should address: between-talker variation within regional dialects and within-talker variation across contexts, and their relationship. First, between and within-talker variability are collapsed and the dichotomy of group informative and talker-informative does not address how within-talker variability is (or is not) captured in the higher-order socio-indexical structure. If the formalization of a talker's accent is the (joint) cue distribution of a given phonetic category, then we have to assume that within-talker variability must be relevant in some way yet remains unspecified. Second, other forms of category structure are treated as separate in the formalization of socio-indexical structure (e.g., dependencies between a cue across two categories). Yet, much work in sociophonetics has emphasized the role of such internal within-category systematicity as an integral facet of socio-indexical structure. In the sections below I will describe the current research on between-talker variation within dialect areas followed by the sources and realizations of within-talker variation. Following these discussions, I will focus on how such individual and group dynamics present potentially distinct distributional challenges. Overall, this section aims to demonstrate the necessity of characterizing the unit (i.e., between or within talkers) and the scope (i.e., grouping level) of analysis to fully characterize the problem (see Tamminga & Wade, 2022 for similar discussion).

7.2.1 Between-Talker Variation

Much work has criticized the homogeneity assumption, suggesting that grouping together several individuals into social groups and assuming, but not confirming, the patterns of the individual, obscures the potential of different linguistic systems between talkers (Bailey, 1973; Bickerton, 1975; Horvath & Sankoff, 1987; Wolfram & Beckett, 2000). Dorian (1994) challenges the homogeneity assumption suggesting that community homogeneity need not

necessarily correlate with linguistic homogeneity. And indeed, several studies that have examined individual variability within social groups have found mixed results about the degree to which individuals demonstrate patterns in line with their social groups. Some studies illustrate talkers are regular and cohesive within their communities (Guy, 1980) while others show individuals may diverge in idiosyncratic ways (Dorian, 1994), or remark upon notable "outliers" or anomalous talkers more generally (Chambers, 2009; Meyerhoff & Walker, 2004; Wolfram & Beckett, 2000; Hall-Lew, 2010).

As noted above, work examining the relationship of individuals to their groups has focused on identifying structured variability across variable forms. Such work has taken multiple methodological approaches, aiming to identify dependencies between different units of analysis (i.e., between or within talkers) and different scopes of analysis (different types of groups), as well as different variable forms under examination. Phonetic dependencies at the individual level have been examined in several ways, including co-occurrence with other categories (or other morphosyntactic features), the linear relationship of acoustic cues across categories sharing phonetic dimensions (e.g., VOT across voiced stops), ranking of systematic patterns (e.g., constraints), and distances between categories (e.g., distance between merged vowels).

Studies that have aimed to examine the relationships among variable forms across individuals within communities aim to uncover patterns of (in)coherence, whereby multiple variables within a community demonstrate correlation in individual talkers' usage (Guy & Hinskens, 2016; Tamminga & Wade, 2022). As defined by Guy and Hinskens (2016:1), "to the extent that linguistic variables systematically covary, they can be characterized as displaying coherence". Several such studies have focused on the relationships among different variables across linguistic levels, such as phonological and morphosyntactic (Erker & Otheguy, 2016; Tsiplakou et al., 2016; van Meel et al., 2016). However, some constrain the units of analysis to the same linguistic level (e.g., phonological variants, Tamminga, 2019). The unit of analysis is most frequently an average of summary statistics, where individual talkers are represented by an average or other measures of central tendency (Tamminga & Wade, 2022). The co-occurrence of variant use is then typically measured linearly (i.e., correlation) across talkers. For example, researchers might represent an individual talker's [æ] and [a] (i.e., the units of analysis) as an average of acoustic cues (F1 and F2), and determine the correlation within a group of individuals

[æ] and [a] position, to determine if individuals who employ a particular variant of [æ] also employ a variant of [a] (see Tamminga & Wade, 2022 and Oushiro, 2016). Such work often examines whether *between*-talker correlations mirror those of community patterns and the degree of dependency between variable forms. As described by Tamminga and Wade (2022), defining the unit and scope of the analysis for identifying such covariation patterns is essential because the wrong scope or unit may lead to erroneously identifying (in)coherence of varieties.

Work in phonetics has also considered dependencies between phonetic forms and abstracted phonological systems, highlighting some degree of uniformity of individuals in terms of the phonetic mapping of phonological targets. In such cases, the dependencies are examined through the linear relationship between low-level phonetic cues and phonological targets, specifically those that share some common dimension (e.g., Chodroff & Wilson, 2022). Chodroff and Wilson (2022) argue the phonology-phonetics interface leads to some degree of phonetic uniformity as targets are shared across phonological primitives (e.g., features such as +/- voice, +/- anterior; see also Fruehwald, 2017; Ménard et al., 2008). Evidence for this comes from the observation that phonetic cues across contrasts are not entirely independent (Allen et al., 2003; Chodroff & Wilson, 2017, 2022; Sonderegger et al., 2020; Theodore et al., 2009). For example, the place of articulation for [s] and [z] are highly correlated within individuals, such that the place of articulation of [s] does not vary independently of [z] within an individual, despite socially meaningful variation linked to [s] (Chodroff & Wilson, 2017, 2022). Similar structured variation can be found in VOT cues to voicing contrasts (Allen et al. 2003; Chodroff & Wilson 2017, 2022; Chodroff et al., 2019; Sonderegger et al., 2020), even when groups differ in their average or central tendency for VOT (Sonderegger et al., 2020). Additional evidence can be found in the covariation of F1 in vowel contrasts sharing phonological height dimensions, such as /eɪ/ and /o/ (Ahn & Chodroff, 2022; Ménard et al., 2008; Oushiro, 2019; Salesky et al., 2020; Schwartz & Lucie, 2019; Watt, 2000).

Such patterns share a commonality with some investigations of coherence in that individual talkers typically represent an average point along cue dimensions across two categories sharing a particular phonological feature (e.g., articulatory dimension). The covariation patterns therein demonstrate that while a particular acoustic cue can overall show high cross-talker variability, they demonstrate a great deal of talker-specificity whereby the

acoustic cues of one category (e.g., /s/) can predict the acoustic cues of another category with shared phonological information (e.g., /z/ by place of articulation). Such talker-specific structure demonstrates high correlations across communities and languages (Chodroff & Wilson, 2022). It is indeed this type of cue-based structured variation that may help listeners discern linguistic contrasts (Chodroff & Wilson, 2018, 2020, 2022; Kleinschmidt & Jaeger, 2015). Evidence of listeners' expectations of structured variation can be found in perceptual learning, whereby listeners extend patterns of VOT from one voiceless stop (e.g., /p/) to another (e.g., /k/) despite exposure to only the former before test.

Such structure has usefully been evidenced to be sensitive to the social group structure, such that groups may vary in terms of their average realization of the acoustic cue, but talkers within each group will demonstrate the same regularity (Sonderegger et al., 2020). There is some dependency between the individual and the group along phonological lines. This type of between-talker variation is perhaps arguably different from the patterns of socio-indexical covariation that may be seen as a result of style, the combination of multiple variants with shared social meaning (see Vaughn & Kendall, 2019). Such factors will be discussed below for completeness, though do not make up the central focus of the types of variability I examine in this dissertation.

7.2.2 Within-Talker Variation

In addition to the degree to which aggregate group patterns are observed between talkers of a group, a great deal of sociolinguistic work has emphasized *within*-talker variability. Another central aspect of the departure from the speech community is the emphasis on individual agency in linguistic behavior, where linguistic features are not inherited from the community, but rather meaning is made "on the ground" (Eckert, 2008, 2012) through a process of linguistic bricolage. Rather than individuals mirroring their community patterns, linguistic bricolage posits that individuals are agentive in the process of linguistic variation drawing on phonetic variation as a symbolic social resource. Support for bricolage is evidenced by speakers' ability to manipulate linguistic form as a function of dynamic social context, producing a range of variation across contexts (e.g., Podesva, 2007). This is not to say speakers have unlimited boundaries on variation, but rather that language users make use of a repertoire of sociolinguistic resources, and

meaning is made during interaction through locally constructed means. Descriptions of within-talker variation are thus a prime example of more distributional properties of categories than other studies that typically examine between-talker variation in terms of averages. This research area points towards complex within-talker variation, which is often reflected in categories that show greater conditioning by social groups, highlighting the need to characterize the interaction more clearly.

Research on individual variation points to the fact that macro-social categories, including geography, are not in themselves indicative of social meaning (Eckert, 2008, 2012; Eckert & Labov, 2017). As noted by Eckert and Labov (2017:470), correlations of gender for example point towards a socially constructed distinction in use, which is an abstraction "over a range of globally constrained but locally constructed practices". However, the shared conditioning between macro social categories and high within-talker variability usefully demonstrates the need to consider the unit of analysis and explication of the nature of input for listeners. For example, categories that are the object of much stylistic variation may have distinct distributional properties within a given talker, including multi-modal distributions (Van Hofwegen, 2013). Multi-modal distributions may also be evident in categories that are undergoing change within a community, as a result of speakers gradually adopting phonetic forms (Fruehwald, 2013, 2017).

Such within-category and within-talker patterns pose some interesting challenges for current conceptualizations of socio-indexical structure. If the formalization of a talker's accent is a cue distribution of a given phonetic category, then we have to assume that such multi-modal distributions must be integrated in some way by listeners. Yet, it's currently underspecified how such within-talker variability fits into socio-indexical structure and how specific categories are identified as informative by listeners. One could presume based on the taxonomy of talker-specific and group-specific categories that within-talker variability would be minimal for group-informative categories, or that the individual and group mirror the same range (i.e., same central tendency and same variance), however, such a component has been unexplored. While this dissertation does not directly address the complexities of the distributional properties associated with social meaning, Chapter 5 begins to examine the interaction between individuals' category distributions relative to their dialect areas.

7.2.3 Taxonomy of Variation

Having established the general areas of debate about individual and group variation, I now turn to describe a taxonomy of variation posited by Guy (1980). Following this, I will revisit the taxonomy in light of distributional properties and current Bayesian models of socio-indexical structure. Guy (1980:12) describes a taxonomy of problems associated with group and individual linguistic variation, focusing again on between-talker variation, suggesting two relevant dimensions: "(a) similarities and differences between individuals (within groups); and (b) similarities and differences between groups". In this taxonomy, variation is hypothesized to appear at the group and individual levels in four different ways, as depicted in Table 2.1.

In the first case, variation is uniformly distributed throughout the community (Table 2.1 #1). For example, /t/-/d/ deletion may comprise such a case where variable realizations of word-final /t/-/d/ occur. In such cases, deletion is uniformly distributed across speakers in the community, despite variable rates of the deleted variant. In the second case, there is variation distinguishing two (or more) different and respectively homogenous groups (e.g., geographic dialects or sociolects; Table 2.1 #2). This case may be exemplified by the canonically identified differences of vocalic variants across regional dialects, where each regional variety is comprised of a distinct variant (i.e., raised /æ/), or differences between social groups such as gender or age. The first and second cases are most aligned with the variation psycholinguistic work assumes when talking about socio-indexical structure at the group level, where groups are internally regular.

The third case is a situation in which different groups demonstrate similar variants but there is a large variety of norms for individuals within the groups. In such cases, group structure arises out of relationships among sociolects, or, at the extreme, individual stratification (Table 2.1 #3). Such cases may occur in cases of idiosyncratic talker variability, or the result of unidentified social formation (i.e., a different group structure). The final case is a situation where combinations of two and three exist together, with individual stylistic variation mixed with inter-group differences Table 2.1. #4). The four different outcomes are not necessarily in opposition to one another, but rather different variables (or categories) may demonstrate different patterns of variation.

Table 2.1 Copied from Guy 1980 (Table 1.2; 12): Types of Structures in Linguistic Variation

Comparing different individuals (within groups)	Comparing Different Groups	
	Similar	Different
Similar	1. Variable rule of uniform force	2. Social or geographic dialects
Different	3. Individually stratified linguistic variation	4. Combinations of 2 and 3, or true free variation

The evidence provided throughout this discussion highlights several complexities in the description and treatment of individuals. Such complexities call for characterization about what it means to have a social group that is informative in so much as they are internally regular and what talker-specific variation specifically references. Several factors must be considered when delineating what it means for an individual to follow the patterns of their social groups including (1) defining the dependency between the group and individual—is it a matter of between or within-talker variability? (2) defining the category relationships (e.g., co-occurrence? Linear relationships? Rates of use?) And (3) how we quantify the unit of analysis and whether it reliably informs the dependency we have defined (e.g., average behavior, distributions, divergence). Below I will provide an updated taxonomy to evaluate the scope and unit of analysis; the primary component of this dissertation addresses between-talker dependencies of regional dialects and individuals, but the taxonomy is meant to illustrate the variable talker behaviors.

7.3 Taxonomy of Variability

The different types of variation have substantial impacts how socio-indexical structure may form over cue distributions across talkers, each having potentially different means by which language users evaluate the speech signal, but which have otherwise remained unaddressed. These outcomes would suggest, at the very least, that language users may learn different models about variation in the world and different patterns based on individual and group dynamics as well as the approach to resolving ambiguity. In Bayesian models of socio-indexical structure,

each individual talkers' "accent" can be formalized as a distribution of acoustic cues and a dialect group's "accent" can be formalized as the distribution of acoustic cues over several talkers (Kleinschmidt, 2019). However, what remains unaddressed is the relationship between individual talkers' distributions of acoustic cues and their respective dialect groups. Walker & Meyerhoff (2004) suggest that variation persists at the level of the individual but that given enough data per talker, individuals are shown to mirror the linguistic conditioning of the speech community of which they are members. Similarly, some scholars argue that within-talker stylistic variation is likely to reflect the range of variation between talkers in a community as correlated with social groups (Bell, 1984; Preston, 1991). Preston (1991) additionally suggests that variation according to social groups is derived from variation in the linguistic context, such that any group level variation is contained within the range of internal constraints on variation.

This fact illustrates the need for more large-scale studies of variation, and the need to look at distributional properties within and across talkers. In light of distributional properties, I revisit Guy's (1980) taxonomy of variation, illustrating the potential patterns of individuals with respect to social groups. A reframing of the types of linguistic variation may be updated to consider how variability is distributed across individuals and communities in terms of their distributional patterns, as reflected in Table 2.2 and Figures 2.7-2.10. Given the original taxonomy concerned sociolinguistic variables more broadly, and not just phonetic categories, the adaptation to phonetic categories will not correspond entirely with Guy's, but extends the core concepts, nonetheless. While this is chiefly concerned with the relationship between individuals and talkers for a single contrast and cue, there is of course additional nuance to how these relationships emerge and are learned. I will not go into detail here about such learning but rather illustrate this as an analytic taxonomy rather than learned and cognitively represented (for discussion of learning of indexicality and speech categories, see e.g., Quam & Creel, 2021).

Table 2.2 Adapted from Guy 1980 (Table 1.2;12): Types of Structures in Linguistic Variation

Individuals	Groups	
	Similar	Different
Similar	1. uniform force; same mean and same variance	2a. Different means; same variance OR 2b. different means and different variance (Social or geographic dialects)
Different	3. Individuals with different means and/or low dispersion (Individually stratified linguistic variation)	4. Combinations of 2 and 3, or true free variation

First, Type 1 (see Figure 2.7) illustrates that over a cue distribution, groups are not differentiated, they show nearly the same central tendency and variance for a cue distribution and individuals are similar to one another. Such an example might be best illustrated by shared linguistic variation, such as vowel length differences preceding voiced and voiceless stops (though, see Tanner et al., 2020 for cross-dialect differences in size of effect). All talkers may apply such patterns uniformly or with some degree of fluctuation which may be the result of noise. Type 1 may be analogous to Ladefoged and Broadbent's (1957) taxonomy of variation termed 'linguistic' variation.

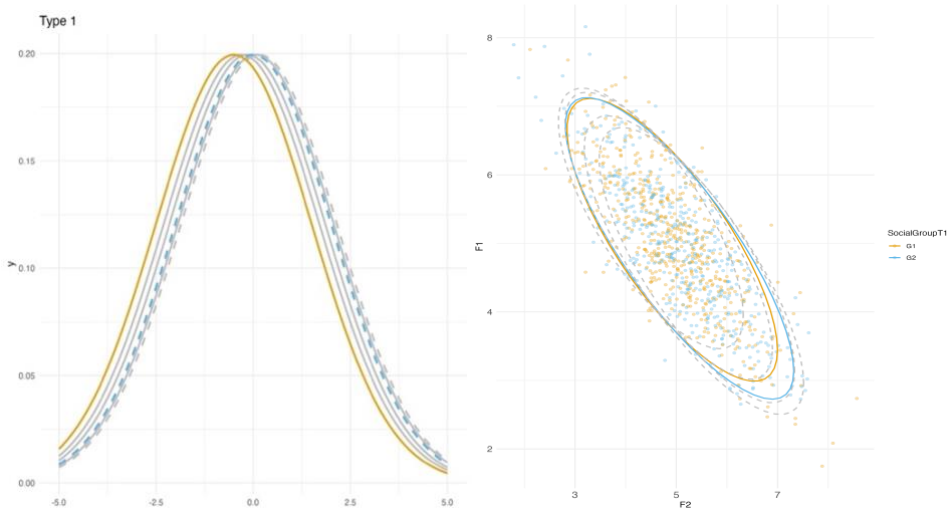

Figure 2.7 Idealized Type 1 pattern, where individual talkers (grey lines) and groups (color and line type) share similar mean and variance (univariate left, multivariate right)

Type 2 may further be broken down into subtypes, as illustrated in Figure 2.8. First, Type 2a (Figure 2.8) illustrates that groups are differentiated by their central tendency but have similar variance, and talkers pattern similarly within but not between groups. Groups are only distinguished by their means and talkers are not meaningfully distinguished within a group. For example, groups of talkers may vary in mean position of /ʃ/, but the variance is similar across groups, and talkers largely align with their social group (e.g., Gunter et al., 2021). In Type 2b (see Figure 2.8) groups are distinguished by their central tendency *and* variance, but talkers pattern similarly within groups. In such a case, we might see that one group has greater variance than the other, and talkers generally mirror the same variance patterns. To be more specific, the degree of within-talker variability is shared across talkers. For example, this pattern may reflect one social group producing variation by phonological context, but for others, it may not. For example, dialect areas may be distinguished by their average /s/ position, and one may further be sensitive to allophonic variation, such as retraction (i.e., more /ʃ/) preceding [tr] contexts, while the other is not, or to a smaller degree (see again Gunter et al., 2021).

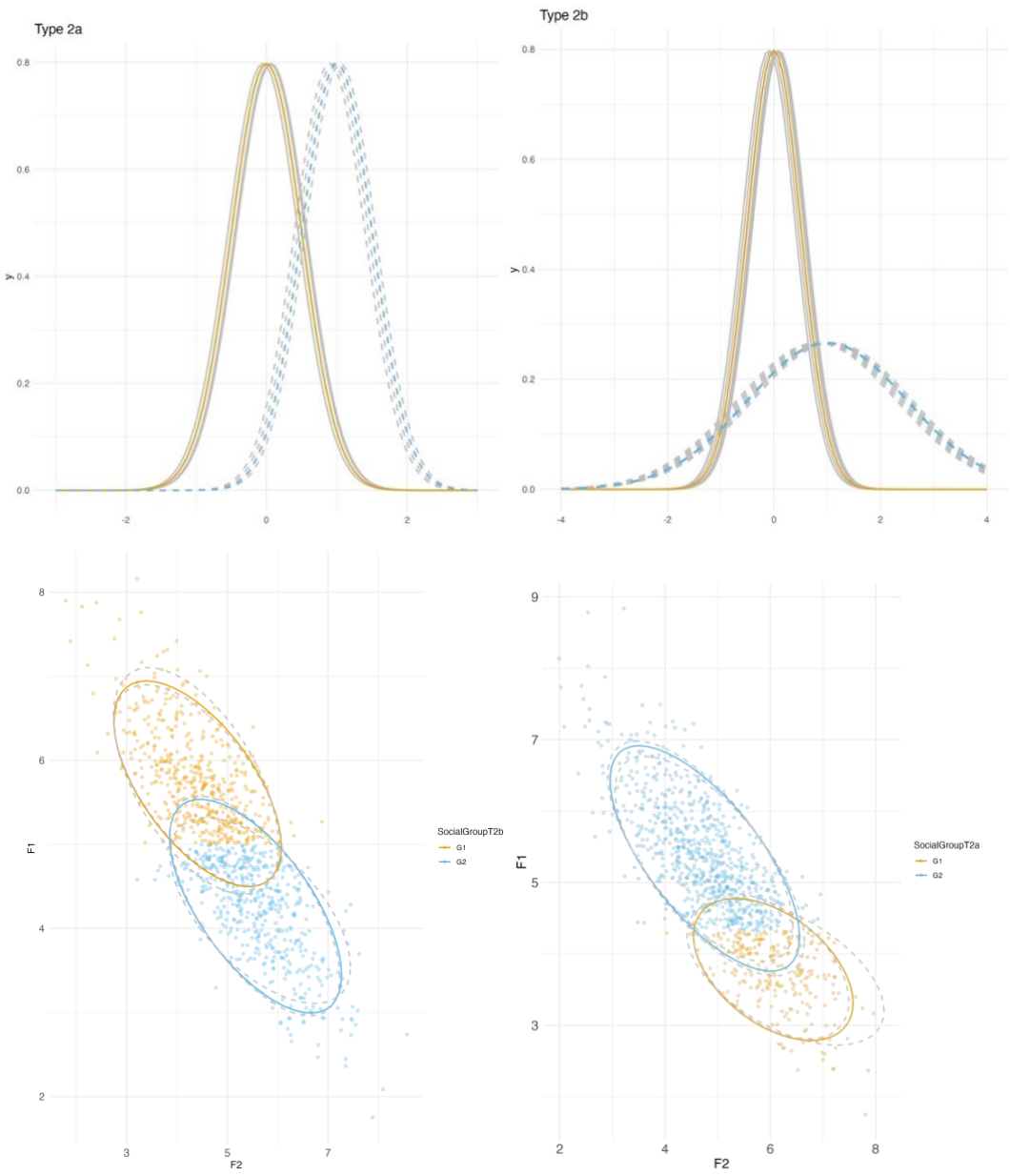

Figure 2.8 Type2a and Type2b, where individual talkers (grey lines) and groups (color and line type) show differences in means and/or variance, but talkers pattern similarly (univariate top, multivariate bottom)

Type 3 is perhaps a case envisioned by current theories of talker-specific variation, where there is no indication of informative social group structure that differentiates individual variability yet there is high cross-talker variability, as depicted in Figure 2.9. In such cases, structure is likely found at a different scope, examining covariation of categories along a shared

phonetic dimension or within category dispersion is low for a given talker showing highly regular forms. An example of this type might be VOT for voiced stops where there is high talker variability across average VOT for talkers, but individuals show cue dependency between contrasts sharing a phonetic dimension, such as voicing (e.g., /b/ and /d/; Chodroff et al., 2015) and talkers have relatively low within category dispersion (Hazan & Baker, 2011). Such a pattern may also reflect the incorrect scope (i.e., social group) defined to capture the socially structured variation or otherwise idiosyncratic variation.

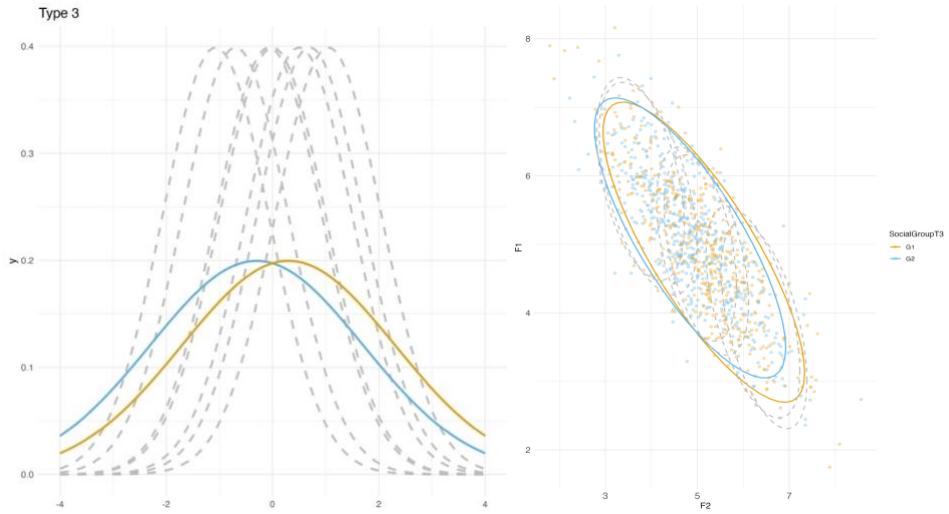

Figure 2.9 Type 3 where individual talkers (grey dashed lines) show differences in means and groups (color) show no differences in means or variance (univariate left, multivariate right).

Type 4 proposes challenges in that there can be any combination of the above factors and may suggest that analysts have defined the wrong scope, unit, or timescale that structures individual variability. On the other hand, it's possible that within-talker variability (i.e., category dispersion) is high, but groups are largely still distinguished and show differences in the aggregate. A possible example is in the case of /s/, whereby gender is predictive of /s/ productions but when you examine individual differences within gender groups talkers show high variability both in terms of between-talker variation (differences in means) and within-talker variability (high dispersion/large variance). As another example, there may be groups such as regional dialects where individuals variably participate in categories affected by regional

shifts (i.e., high between-talker variability) or within-talker variability is high. As an illustration of the distributional properties, see Figure 2.10.

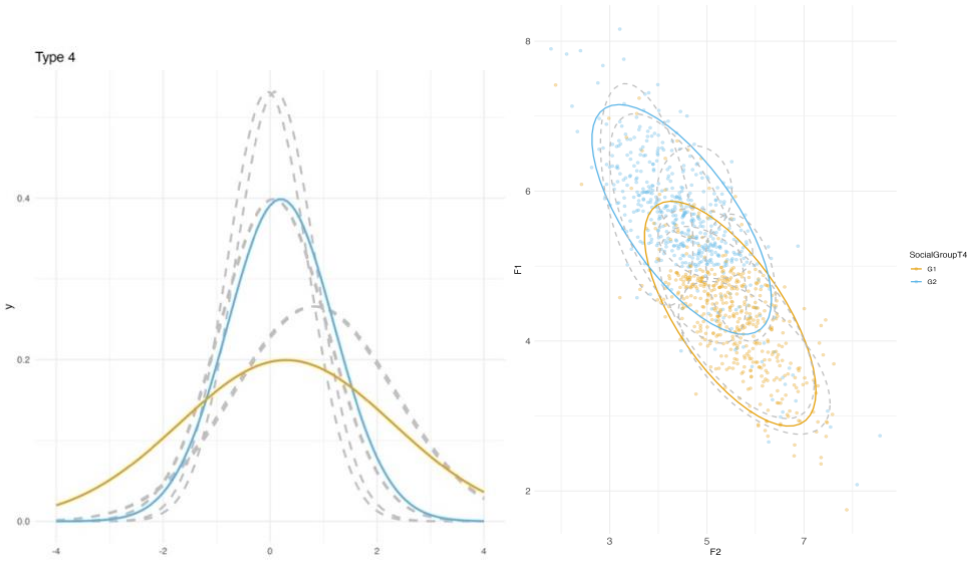

Figure 2.10 Type 4 individuals and groups show variable realizations in mean and variance, with no clear grouping structure (univariate left, multivariate right).

The distinctions between each type in the taxonomy are largely idealized, and researcher observations of such patterns may demonstrate differences in the scope of the analysis, or if using macro social categories, incorrectly define the group. Given these various distinctions, however, and the emphasis on internal linguistic structure in relation to socio-indexical structure, theories would benefit from characterizing the nature of socio-indexical patterns and the consequences of variability and listeners' beliefs. To characterize listener-oriented behavior, Bayesian models of socio-indexical structure must attempt to characterize the assumptions of the prior. As noted above, Type 2 largely reflects modern discussions of how listeners learn socio-indexical structure in Bayesian models: largely homogenous groups provide cues to socio-indexical information and language users learn this statistical relationship. In this case, the phonetic space is readily partitioned based on the regularity of statistical differences between groups, where language users within those groups are consistently aligned. Language users then build causal models for perceptual processing to align with these beliefs, such that contrasts that show regularly conditioned group variability would likely result in more flexibility in adaptation

as well as greater generalization of novel patterns due to the likelihood of such variability belonging to many individuals of a social group, rather than an idiosyncratic pattern (Kleinschmidt, 2019).

Type 3 predominately reflects the systematic between-talker variability discussed by Kleinschmidt (2016, 2019; see also Kleinschmidt & Jaeger, 2015). With extreme individual stratification, there should be little benefit for language users to infer any social grouping—that is the "community" is merely an imposed abstract grouping of individuals by the analyst, but by which language users would be unlikely to construct from experience with the individual talkers. However, individual talkers' distributional patterns still aid in predicting and parsing the speech signal in talker-specific ways. In such cases, prior experience with other talkers provides little benefit and listeners are more likely to learn quickly and more robustly when encountering novel variation. In addition, low category dispersion may provide greater certainty in a talker's cue distribution resulting in faster and easier categorization while greater dispersion results in less stable identification responses (Clayards et al., 2008; Drouin et al., 2016; Newman et al., 2001; Theodore & Monto, 2019). Indeed, perceptual learning literature provides evidence of adaptation to individual talkers' VOT and subsequent generalization to shared contrasts within a talker (Kraljic & Samuel, 2006; Kraljic et al., 2008; Lai, 2021; Munson, 2011), suggesting such regularity may be leveraged by listeners.

In the case of Type 4, however, there is very little to be said about how language users learn the relationships between socio-indexicality and variation of these types. Type 4 could potentially be divided into two types of variation with distinct ramifications for listener-oriented behaviors. In what I'll call Type 4A, the community may demonstrate more hierarchical or nested organization among sociolects, such that the larger group patterns are only a sum of the other sociolects in the community, so individual adherence may be more likely observed in a mediated social grouping (i.e., a different analytic scope). If this is the case, then current perception models should be able to accommodate this structure in a similar way to Type 1 and 2, so long as the sub-groups are regular and homogenous. Given the emphasis on bottom-up grouping of related talkers in current models, this is of course plausible. However, given that much regional variation may demonstrate such patterns, it's not clear how inferences about socio-indexical factors guide speech perception. Namely, at what level of such a hierarchical

organization do listeners make inferences about the causality of variable input? This is an open question, and one I will discuss in more detail in subsequent chapters, though will not fully resolve.

Type 4b is, potentially, the most complicated situation for our models of probabilistic learning where there are regular group differences in production but individuals' adherence to these group norms is the object of much stylistic variation and within-category dispersion is high. In this case, it's unclear what language users would build as their beliefs: that the individual is in flux and variation will always be, to some level, dependent on context and the individual or whether there is some abstraction that occurs over the exemplars of individuals across different contexts who all belong to the same group. Do language users form expectations based on the probabilistic cues associated with individuals in contexts, and disregard "group" information? Or do language users maintain bivalence of both the group norms and the lower-level stylistic information? The answer to this may depend both on the language users' experience with this type of variability, whether they are a member of the same community, and of course the task. If the speech community does function because users have shared evaluations of linguistic norms, then it would make sense that these more fine-grained stylistic choices would be tracked alongside larger community norms. However, if language users do not come from the same local community, then they may build causal models that only account for the larger community-level patterns that are more statistically regular in their experience. Both explanations are plausible in Bayesian models, and as noted by Kleinschmidt (2019:7), "group-conditioned cue distributions reflect the starting point for **talker- or situation-specific** distributional learning" [emphasis added]. However, the nuances of how group-conditioned cue distributions prove to be informative with such inter- and intra- talker variation remains to be seen. Given the nature of the data in this dissertation, I will not be able to fully speak to this type of stylistic variation but examining distributional properties in vowels may act as a first step in addressing these questions.

Overall, this may point to the fact that listeners may learn patterns of phonetic variation in line with different models. While a process of linguistic bricolage could facilitate maximal talker recognition, a process more akin to maximal community structure could ideally provide less need for adaptation and learning in processing. The more individually specific patterns of

socio-indexical structure may potentially facilitate processes of talker recognition of inter-talker differences in phonetic realizations of contrasts in idiosyncratic ways. While variability that is systematically structured at the group level may facilitate talker recognition across a wider range of talkers within a particular group (Docherty & Foulkes, 2014; Kleinschmidt, 2019). Vowels provide an interesting test case for identifying structured phonetic variation. On the one hand, vowels demonstrate robust dialectal differences, and listeners have some degree of awareness of vocalic variation across groups. On the other hand, they are a rich source of stylistic variation, where talkers seem to draw on individual categories for meaning making (Eckert, 2012; Eckert & Labov, 2017) and provide extensive talker information from spectral cues (e.g., Kleinschmidt, 2019).

8 Conclusion

In this Chapter I have attempted to lay the groundwork for components of a theory of socio-indexical structure, drawing on work from across different domains of linguistic inquiry, emphasizing the role of the listener. Drawing from work in sociophonetics it is clear that structured variation occurs as a result of both internal linguistic constraints and external social factors. Despite the variability within social groups and individuals, listeners demonstrate a great deal of shared latent knowledge, especially with respect to regional dialects, that shapes their perception of linguistic categories. Current theories in psycholinguistics and sociophonetics suggest that listeners learn socially meaningful variation by tracking the statistical regularities of the speech signal across talkers and their social groups which is supported through distributional learning mechanisms. When listeners engage in linguistic comprehension, they then use prior knowledge about cue distributions and the social correlates that condition them to predict the incoming speech signal, adapt to novel talkers and novel forms, and generalization to similar talkers. While such a system is supported by much work in perceptual learning, the specifics of what listeners track and when they generalize are still unclear. Work in sociolinguistics provides a general lens through which we can begin to characterize such components allowing us to integrate various literatures to build a comprehensive model of socio-indexical variation. To move towards a comprehensive model, in this dissertation, I hope to examine different

conceptualizations of the prior that make up listeners' input and test one such prediction in perceptual learning.

The primary themes I hope to address in the chapters to follow include: 1) how we characterize individuals and their relationship to regional dialect; 2) how we characterize the prior which comprises the input for listeners' representations and inferences; and 3) how do such listener inferences predict perceptual learning behavior. To address these questions, I examine vowel categories as a paramount example of structured variation, but which has otherwise been underexplored in the domain of perceptual learning. Using corpus data, I attempt to analyze different baseline experiences with regional and talker variability in American English assuming raw frequency distributions as the input (Chapter 4). Following this, I revisit core concepts in sociolinguistics about the internal properties of vowels and cue-specific tendencies to elucidate how such specificity can work in tandem with socio-indexical structure (Chapter 5). From there, I provide an example of the types of questions that can be drawn from corpus data, testing asymmetries between two vowels that share critical differences in group and talker properties in a perceptual learning experiment (Chapter 6). Following these analytic chapters, I will return to the themes and questions raised in this chapter in the discussion.

CHAPTER 3:
CORPUS DATA & PROCESSING

1 Introduction

The first part of this dissertation is comprised of several corpus analyses aimed at assessing both the internal and socio-indexical structure of variability across talkers and dialect areas in American English. This Chapter describes the data and related pre-processing that feeds into the analyses in Chapters 4 and 5 and provides initial breakdowns by corpora. Within each chapter, relevant analytic choices and subsets of data will be detailed in line for each analysis, which are drawn from the overall dataset outlined in this Chapter. In the following sections, I will describe the source of the data and representativity for the goals of this dissertation (Section 2), specify detailed information about the original source corpora and speaker demographics (Section 3), the automated acoustic measurement (Section 4), and the post-processing procedures (Section 5). The datasets are drawn from an open-source repository curated by the Speech Across Dialects of English (SPADE) project (Stuart-Smith et al., 2019; Mielke et al., 2019). Thus, Sections 3 and 4 represent the collaborative efforts of many researchers who contributed to the project. While I did not own the data collection or measurement, I was part of the SPADE project team, and knowledge about the project in these sections stems from this collaboration. Section 5 (along with Chapters 4-5) specifically describes my individual efforts toward data validation and processing.

2 Data & Representativity

The data for this dissertation come from the SPeech Across Dialects of English (SPADE) project (Stuart-Smith et al., 2019; Mielke et al., 2019). The SPADE project developed and applied software to pre-existing corpora for large-scale speech analysis. The data are composed of existing private corpora collected by various researchers along with publicly available corpora. The SPADE project aims to make data accessible to researchers either directly (e.g., directly downloading datasets) or indirectly in the form of acoustic data that can be queried (but not accessed directly) to maintain participant privacy (Sonderegger & Stuart-Smith, 2022).

The datasets used in this dissertation were selected from a larger repository containing automatically extracted acoustic measurements using SPADE tooling. Automatic extraction was carried out by team researchers using the Integrated Speech Corpus Analysis (ISCAN) software developed as part of the SPADE project. The subset of datasets used in this dissertation was selected to represent American English broadly and are drawn from nine source corpora, which together cover seven geographic dialect regions of the U.S.: North, South, West, Northeast, Midatlantic, Midland, NYC (see Figure 3.1 for overview and Section 3.1-3.8 for details about the source corpora). Each dataset consists of automatically extracted static formant measures (F1, F2, F3) from one-third of the duration of the vowel, as well as vowel duration drawn from the phone-aligned transcripts.

The corpora comprise a range of speech styles (from read speech and word lists to sociolinguistic interviews), demographic backgrounds (age, gender, ethnicity), time of recording, and recording equipment (more details in Section 3 below). As such, the data represent a diverse perspective of American English and regional variation and was not curated to represent the most divergent or representative speech from each region. I believe this has several advantages, including a more ecologically valid representation of the speech listeners are likely to encounter on a regular basis. Such ecological validity provides a robust sample for validation of prior studies' findings and better positions researchers to form predictions and hypotheses around listener expectations based on prior experience. Given that listeners' expectations are drawn from a noisy and variable experience, these data provides an opportunity to probe precisely how such noisy and variable experiences are systematically organized in production and characterize listeners' expectations in perception.

Of course, the noisier and less controlled data also pose some drawbacks, including less clearly demarcated patterns and potential sensitivity to noise in statistical analysis. Similarly, the data does not necessarily represent the most divergent or representative speech for a given dialect area. To counter these challenges, I draw on several simulations and statistical methods to capture the most well-rounded perspective of the data as possible, including analyzing specific subsets and levels of socio-indexical organization (from individuals to larger dialect areas). Further, the intra-regional variability is of theoretical import to both how variable speech is structured in production and how listeners make sense of such variability. In addition to the

theoretical import, such statistical experimentation allows for a better understanding of the patterns above and beyond the potential impact of noise. For these theoretical and methodological reasons, the range of variation in the dataset is a strength of the analysis.

3 Corpora & Speakers

Before turning to the brief overviews of each of the corpora and the demographic breakdown of speakers (Sections 3.1-3.8), I will briefly describe the processing and aggregation of speaker metadata across the datasets. Speaker metadata was retained from the source corpora, with the exception of dialect area and ethnicity, which were aggregated for analytic purposes (dialect) and ease of comprehension (ethnicity). Fine-grained metadata about each individual's location of residence (or length of residence) was not available across all corpora. Accordingly, using the available geographic information about speakers, dialect regions were standardized around dialect areas delineated in the Atlas of North American English (ANAE; Labov et al., 2006). Figure 3.1 and Table 3.1 provide the overall geographic spread and relevant corpora across the datasets. For one of the corpora (Switchboard, see Section 3.1) broad dialect areas were the only available metadata regarding speakers' regional background limiting information for the determination of boundaries of dialect area. For this case, speakers' regional affiliation was preserved from the source data, corresponding (roughly) to the dialect areas of the ANAE, with the exception of the North and South Midland, which were collapsed into one 'Midland' region.

Comparably, ethnicity and race demographics varied in collection and reporting across the source corpora. While these demographic details are not part of the social factors examined explicitly in this dissertation, they are nonetheless represented in the speech; Table 3.2 reports the number of speakers within macro racial and ethnic categories, and individual corpora speaker distributions are reported in Sections 3.1-3.8 below. Given the varied reporting, the macro categories represent an analytic aggregation of speakers and not necessarily their self-reported racial or ethnic identity. For example, the label 'Asian or Asian-American' includes a variety of identities provided by participants including 'Asian-American' and 'Taiwanese'. Likewise, 'multiracial' aggregates speakers who reported bi- or multi-racial identities, across variable combinations. Unfortunately, the Switchboard corpus did not have racial or ethnic identity of the

participants published, thus the "unknown" category reflects this pattern. The 'other' group represents only speakers where 'other' was originally specified in the metadata.

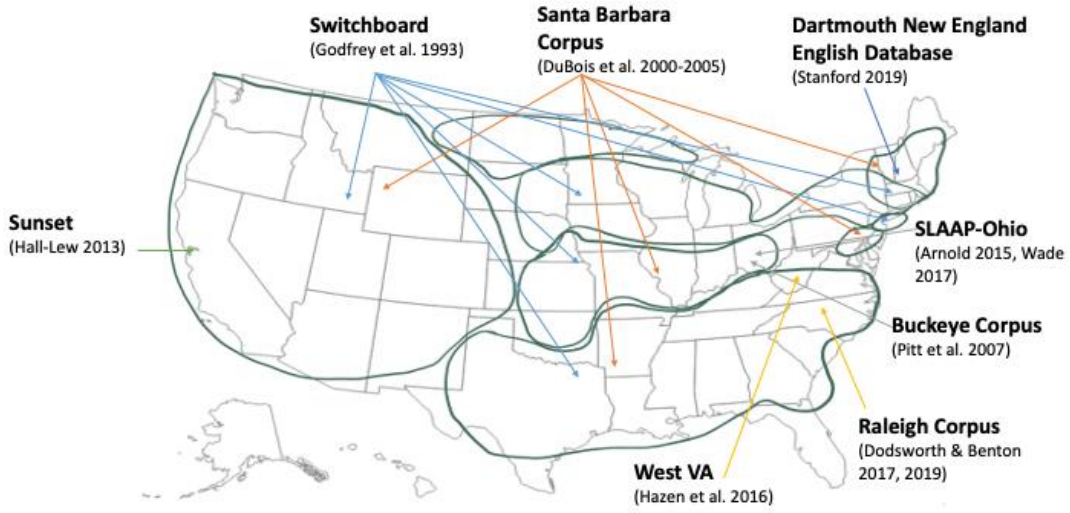

Figure 3.1 The corpora represented in this dissertation and the respective geographic regions represented by talkers in the corpus.

Table 3.1 Total speakers grouped by gender and dialect area across all corpora.

Dialect	Female (N)	Male (N)	Total
Midatlantic	10	7	17
Midland	153	156	309
North	21	37	58
Northeast	109	109	218
NYC	8	11	19
South	107	109	216
West	62	58	120
Total	**470**	**487**	**957**

Table 3.2 Total speakers by broad racial and ethnic groups. Groups are researcher aggregated based on original ethnicity reported across corpora. The unknown category represents Switchboard data, for which ethnicity was not reported.

Ethnicity	N
African American or Black	27
Asian or Asian-American	11
Native American	3
Caribbean American	2
Jewish	3
Latinx/Hispanic	12
Multi-racial	8
other	5
unknown	339
white	546
Total	957

3.1 Switchboard (Godfrey & Holliman, 1993)

Originally collected by Texas Instruments, this corpus contains about 2,400 telephone conversations recorded 1990-1991. A total of 543 (241F, 302M) participants ranging across geographic location in the U.S. A subset of the data are represented here, with a total of 339 speakers (152F, 1867M). The data was forced aligned by the SPADE team using MFA (McAuliffe et al., 2017). Speaker metadata includes gender and dialect area, as coded by the original researchers. Speakers were kept in their original dialect areas (see Table 3.3), except in the North and South Midland, which were combined into a larger 'Midland' category for this dissertation. Ethnicity and racial identity were not provided for the Switchboard data and were categorized as 'unknown'.

Table 3.3 Unique number of speakers by gender and dialect area, as originally coded in Switchboard (i.e., South midland and North Midland are kept separate)

Dialect	Female (N)	Male (N)	Total
North	21	37	58
North midland	17	30	47

Table 3.3, Continued

Northeast	10	8	18
NYC	8	11	19
South	16	22	38
South midland	59	50	109
West	21	29	50
Total	152	187	339

3.2 Santa Barbara (Du Bois et al., 2000-2005)

The Santa Barbara corpus was compiled by a group of researchers in the Linguistics Department at the University of Santa Barbara, directed by Dr. John W. Du Bois. This corpus contains about 200 (~ 99F, 79M, 32 'unknown') speakers of conversational or naturally occurring speech recorded in the late 1990s through the early 2000s. A subset of the total speakers are represented in this dataset, with a total of 135 speakers (75F, 60M) across five regional dialects. The corpus has a dataset of mixed ethnicities, of which the majority are white (N = 113), followed by Latinx (N = 11), Native American (N = 3), African American or Black (N = 2), multi-racial (N = 1) and other (N = 5). The corpus was aligned as part of the SPADE project using MFA (McAuliffe et al., 2017). Speaker metadata include gender, dialect state, age, hometown, ethnicity, and information on education and occupation. Speakers were categorized into dialect areas based on their dialect state information, as summarized in Table 3.4.

Table 3.4 Unique number of speakers by gender and dialect area, categorized by dialect state information from the original corpus.

Dialect	**Female (N)**	**Male (N)**	**Total**
Midatlantic	10	7	17
Midland	23	21	44
Northeast	8	7	15
South	9	7	16
West	25	18	43
Total	**75**	**60**	**135**

3.3 Sunset Corpus (Hall-Lew, 2013)

The Sunset Corpus was originally collected by Dr. Lauren Hall-Lew in the Sunset district of San Francisco (Hall-Lew, 2013). The corpus contains conversational sociolinguistic interviews of 28 participants, of which one speaker is not in the current dataset (16F; 11M). The speakers are European-American (N = 11) and Chinese American (N = 16) speakers as reflection of the demographics of the Sunset neighborhood (recorded in 2008-2009). The corpus was aligned by the SPADE team using MFA (McAuliffe et al., 2017). Speaker metadata include pseudonym, gender, year of birth, interview year, ethnicity, and heritage language. All speakers in this region were coded into the 'West' dialect region for this dissertation.

3.4 West Virginia (Hazen et al., 2016; Hazen, 2018)

The West Virginia corpus was collected as part of the West Virginia Dialect Project led by Dr. Kirk Hazen to document language variation in the speech of Western Virginians and educate the broader public. This corpus contains sociolinguistic interviews of 61 (31F, 30M) speakers as part of the West Virginia Dialect Project. The speakers are white (N = 54) or African American or Black (N = 7). The data were aligned by the original researcher using FAVE (Rosenfelder et al., 2015). Speaker metadata include birth year, sex, ethnicity, hometown, education, rural/non-rural, class, region, age group. The speakers were all categorized into the 'South' dialect region for this dissertation.

3.5 Raleigh (Dodsworth & Benton, 2017)

The Raleigh corpus was originally collected by Dr. Robin Dodsworth and represents sociolinguistic interviews of speakers from Raleigh, North Carolina (recorded 2008-2017). The dataset available from SPADE contains only a subset of the sociolinguistic interviews for a total of 101 (51F, 50M) white speakers. Recordings were aligned by the original researchers using P2FA (Yuan & Liberman, 2008). Speaker metadata includes ethnicity, year of birth, and gender. The speakers were all categorized into the 'South' dialect region for this dissertation.

3.6 Dartmouth New England English Database (DNEED; Stanford, 2019)

The DNEED was collected as part of the Dartmouth New England English research project, led by Dr. Jim Stanford in collaboration with Dartmouth undergraduates over the span of 8 years (recordings from 2010-2017). The data represent speech from across all six New England states. While the original corpus contains both interview and read speech, the SPADE data used in this dissertation consists only of the read speech of 185 (91F, 94M) participants, who were asked to read 12 sentences each (roughly 16 hours of speech). Phone-level segmentation was done during corpus creation using FAVE (Rosenfelder et al., 2015). Speakers are from a variety of ethnic backgrounds, with the majority being white (N = 163), followed by African American or Black (N = 16), Caribbean American (N = 2), Jewish (N = 2), Latinx (N = 1), or multiracial (N = 1). Speaker metadata includes gender, origin, hometown, birth year, education, place, ethnicity, occupation, state, child latitude, child longitude. All speakers were categorized into the 'Northeast' dialect region for this dissertation.

3.7 SLAAP-Ohio (Arnold, 2015; Thomas, 2019; Wade, 2017)

The Sociolinguistic Archive and Analysis Project (SLAAP; Kendall, 2007) is an interactive web-based archive of sociolinguistic recordings which includes a suite of corpus and phonetic analysis tools. The Ohio data here represent a subset of data available in SLAAP and 1 of 6 SLAAP datasets available in SPADE. The Ohio data were originally collected by Dr. Erik Thomas beginning in 1993, and with additional collection by Dr. Lacey Wade (spanning to 2015). A total of 69 speakers (34F, 35M), from various places in Ohio are represented in the SLAAP-Ohio data. The majority of the speakers are white (N = 66) followed by African American or Black (N = 2) and multiracial (N =1) identities. The data were forced-aligned as part of the SPADE project using MFA. Speaker metadata include sex, year of birth, ethnicity, and locality. Speakers were categorized as the 'Midland' dialect area for this dissertation.

3.8 The Buckeye Corpus (Pitt et al., 2007)

The Buckeye Corpus was originally collected by researchers at Ohio State University. The corpus contains conversational interview speech data of 40 (20F, 20M) white speakers from

Columbus, Ohio recorded in 2000. The phone alignment segmentation was corrected as part of the collection process. The original segmentation was phonetically aligned by the original researchers and was converted to phonological labels as part of the SPADE project. Gender and age are available as speaker metadata. Ethnicity was added for this dissertation based on details provided on the original corpus webpage(s) but was not included in the original dataset. The speakers in this corpus are categorized into the 'Midland' dialect region for this dissertation.

4 ISCAN & Automatic Extraction

ISCAN is an open-source software developed by the SPADE project for analysis of spoken corpora varying in format and size by enabling automated acoustic phonetic extraction (McAuliffe et al., 2019; Mielke et al., 2019). Such software advances the phonetic analysis of corpora by enabling replication by replicating acoustic extraction methodology across corpora. For formant extraction, ISCAN uses a bootstrapping measurement under different LPC coefficients at one-third of the vowel's duration. Each of the measures is compared against a prototype that is parametrized by mean and covariance matrix of relevant formants and bandwidth. An additional algorithmic step identified potential errors in formant measures and performed reanalysis to drop individual formants and retain the next highest formant measure. The selection of the final measurement is based on the smallest distance from the prototype (as measured using Mahalanobis distance). Following automatic extraction, Mielke et al. (2019) preformed accuracy validation by comparing the datasets to manually computed measurements of vowels. For full details regarding the automatic measurements and validation, see Mielke et al. (2019). The SPADE team did the automatic extraction of vowel data using ISCAN and the data for this dissertation were retrieved from the OSF repository post-extraction and pre-processing.`

5 Data Post-Processing

The cues relevant to vowel quality used in this dissertation are the first and second formants (F1, F2 respectively). As noted above, F1 and F2 were automatically measured at 1/3 the vowels duration, where duration is based on the aligned intervals of the transcribed audio. Other cues and dynamic changes in F1 and F2 are undoubtedly used by listeners during linguistic

categorization and in social identification of speakers, however, F1 and F2 represent the predominant cues associated with vowel quality (Hillenbrand et al., 1995). This selection of these cues for this dissertation is not meant to suggest that these are the only cues that matter, and structured variation does not occur across other cue dimensions. Rather, F1 and F2 represent a worthy starting point by which we can begin to explicate socio-indexical structure across vowel categories as it remains the predominant means of describing vowels and vocalic variation across the field.

After collection of the original datasets as provided by the SPADE team, all data were subject to the same post-processing to create parity and consistency across the data. SPADE originally provides vowel category labels based on MFA extracted arpabet labels, original corpora transcription conventions, or Unisyn Lexicon (Fitt, 2000) specific label conventions that broadly align with Wells's (1982) lexical classes. All labeling conventions were broadly grouped into single phonemic vowel labels consistent with American English. A stop word list was applied to the datasets which removed high frequency function words (e.g., *he, they, am, an,* etc.), taken from the Dartmouth Linguistic Automation suite (Reddy & Stanford, 2015). Common lexical items with variation that is non-generalizable to a class (e.g., *tomato, basil, either, neither*) were removed (N = 4898) due to the irregular phonetic forms used to transcribe such items. Further, rhotic items transcribed as syllabic or r-colored (e.g., ER) were removed for analysis (N = 19767). Finally, only stressed vowels were retained for the analysis, either primary or secondary stress, eliminating vowels that were potentially unstressed and reduced.

F1 and F2 values were normalized using Lobanov normalization (Lobanov, 1971), which effectively z-scores the Hz values by participant, scaling and centering the vowel space. Lobanov normalization has been demonstrated to preserve dialectal differences while removing gross physiological differences, which are often correlated with gender (Kohn & Farrington, 2012; Thomas & Kendall, 2007). Contrastingly, gross physiological differences are preserved in un-normalized (and in Bark transformed) space and may obscure differences across regional dialects, of which is the primary factor of interest in this dissertation. Lobanov is not unique in this respect and is not meant to represent a model for listener normalization and rather represents an analytic choice for data processing. Similarly, it is not to say that gender is not an informative component of listeners' perceptual processing, but rather as a means of reducing the influence of

gender on raw formant values to facilitate the identification of regional patterns. Further, it allows for better cross-discipline and cross-analytic perspectives, as it is the predominate method used in sociophonetic studies of regional dialect variation (Kendall & Fridland, 2021) and used in current work in speech processing that this dissertation draws from (e.g., Kleinschmidt, 2019). Thus, it provides greater reproducibility and interpretation of the findings of previous work in addition to its analytic benefits as a normalization method.

Post normalization, tokens were removed that had a duration of less than 60msec (N = 59,255). Tokens greater than 2.5 standard deviations on a by-speaker basis were removed (N = 9,011). The final dataset totaled 586,945 tokens across 13 vowel categories. For the analyses in the following chapters, I chose to only look at 11 monophthongs and /aɪ/, given its prevalence in vowel shifts (see Chapter 2), which provided a total of 556,946 tokens. After which, rows were removed where the speaker's vowel category had fewer than 5 tokens (i.e., speaker001 /aɪ/ < 5 tokens that speaker's /aɪ/ category was removed); a total of 3,898 items were removed from the analysis and 17 unique speakers leaving a total of 583,047 tokens across 940 speakers.

6 Conclusion

Overall, the data present a unique opportunity to validate assumptions about talker variability in American English. The wide range of styles and talkers simulate the broader range of variability that listeners contend with for speech-processing tasks and allow for varied simulations to aid in the characterization of listener experiences and resulting behavior. Similarly, drawing on large-scale corpus phonetics can provide researchers with insights into phonetic variability and questions about internal systematicity. These points will be illustrated in more detail in the following chapters.

CHAPTER 4:
PRIOR EXPERIENCE & SOCIO-INDEXICAL GRANULARITY

1 Introduction

As described in Chapter 2, recently linguists' understanding of variability has begun to converge on the utility of structured variability in speech perception processes, with psycholinguistic theories building variability into theoretical frameworks (Bent & Holt, 2017; Clayards et al., 2008; McMurray & Jongman, 2011; Jongman & McMurray, 2017; Kleinschmidt & Jaeger, 2015; Kleinschmidt, 2019; Kraljic et al., 2008). Drawing on Bayesian models of speech perception, recent iterations of ideal adapter theories have progressed to hypothesize the role of socio-indexical structure in online speech processing (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). These models posit that listeners track the statistical contingencies between social factors and phonetic variability when it is informative to speech processing. Informativity is defined by talkers grouping into regular and coherent socio-indexical groups which can partition the variable phonetic space. A key formalization of ideal adapter models suggests that talkers'—and by extension social groups'—accents are a probability distribution for a given cue and category mapping from which listeners learn talker- and group-specific characteristics (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015). The learned mapping of phonetic variability to socio-indexical factors consequently informs initial speech perception and acts as the starting point for further adaptation to novel talkers. Adaptation depends on listeners' a priori experience with cross-talker variability. As such, research should be cognizant of how prior experience may vary across individuals and our treatment of what constitutes the baseline experiences (i.e., the prior) listeners have.

This chapter attempts to elucidate such descriptions of listeners' priors to inform predictions regarding listener behavior in adaptation. The input for learning, according to Kleinschmidt (2019), may be described as a generalized property, whereby the (joint) cue distribution for a given contrast (e.g., vowels, stops, etc.) provides input for both linguistic and social inference. I will label this perspective as the *holistic* hypothesis, whereby vowels overall are informative of cross-talker variability and listeners' perceptual learning stems from this

generalized description. Using different analytic conceptualizations of socio-indexical structure, this chapter challenges the *holistic* hypothesis. By examining different degrees of granularity in socio-indexical structure, we can begin to hypothesize how different analytic levels predict diverse listener behaviors. The analysis in this chapter extends recent work by Kleinschmidt (2019) drawing on and validating information theoretic measures of *informativity* (see Section 3) for evaluating socio-indexical structure in speech. As such, I make some of the same simplifying assumptions as Kleinschmidt (2019), including formalizing individual talkers' and dialect areas' varieties according to the joint cue distributions (F1xF2) of vowel categories. However, this chapter seeks to examine three additional research themes to augment and nuance the findings from Kleinschmidt (2019):

- How do different ranges of previous experience change how informative socio-indexical factors are of individual vowel categories?
- How does the organization and granularity of socio-indexical structure provide alternate predictions?
- To what extent do the distributions of individuals map to the distributions of their groups? How do individuals diverge from their regional groups?

Over three analyses below (Section 4) I evaluate how socio-indexical factors condition variability across simulations of different baseline experiences using a large-scale dataset of vowel measurements across American English (see Chapter 3). The simulations aim to validate the extent to which different prior experiences provide comparable emergence of socio-indexically conditioned variation across the vowel space. In addition, I extend the methodology to analyze different organizations of socio-indexical structure to determine how different degrees of specificity or generalization across vowels and talkers provide further insight into socially structured variation across the vowel space. As a subset of this, I evaluate how individual talkers' distributional properties (i.e., within-talker variability) align with their dialect areas' distributions.

These questions speak to broader themes in this dissertation (as described in Chapters 1-2), including how individuals align, or not, with their dialect areas, and inquiries about the

degree of specificity in socio-indexical structure for listeners. In addition, the analyses in this chapter provide insight into socio-indexical structure across regional varieties in production, adding an under-described perspective to ongoing work on regional vowel patterns. The findings of this chapter highlight several broad takeaways. First, counter to the holistic hypothesis, vowel categories generally show asymmetrical conditioning, where some vowel categories are more strongly conditioned on dialect, while others are more robustly conditioned on individual talkers and lack dialect conditioning. Across the several simulations, two vowel categories emerge most frequently that reveal this asymmetry: /eɪ/ as dialectally conditioned and /ʊ/ as talker conditioned. Second, within dialectally conditioned vowel categories, I observe some degree of low-level phonetic uniformity across talkers. Finally, vowel categories that are most commonly described across regional vowel shifts are more likely to demonstrate greater between-talker variability both within and across regions.

In the following section I will review Kleinschmidt (2019) for context and methodological points (Section 2.1). Following this, I will outline current literature and hypotheses about dialect variation in production, focusing specifically on the tension between individuals and community patterns (Section 2.2). Following the background, I will discuss the methodology (Section 3) before moving on to the three primary analyses of this chapter (Section 4). Finally, I will summarize and discuss the implications of the results in more detail and conclude (Section 5).

2 Background

2.1 Informativity

Kleinschmidt (2019) specifically tackles the problem of the possible parameters of socio-indexical structure that are available to listeners in his computational-level theory of the ideal adapter model. The computational model aims to distinguish between the mere existence of socio-indexical differences and which of these differences are meaningful or worth tracking for speech perception. Within Kleinschmidt's (2019) model of the ideal adapter, socio-indexical structure is formalized as the informativity of a socio-indexical factor for speech processing as conditioned on a contrast's cue distributions. Informativity is quantified using Kullback-Leibler

(KL) divergence, an information theoretic measure that quantifies how similar (or divergent) two probability distributions are (Kullback & Leibler, 1951; see Section 3 below for specific details). In Kleinschmidt's model, he compares the cue distributions of socio-indexical factors to a marginal distribution reference group made up of American English more broadly. As such, the model assumes that the reference group is representative of American English and the starting point from which listeners evaluate the extent to which a socio-indexical group may provide information for speech perception. When a socio-indexical factor is highly divergent from the American English baseline, it suggests there is more information to be gained about cue distributions from knowing the more specific social grouping (e.g., gender) and thus more likely to be tracked by listeners. Thus, informativity might uncover reasonable listener expectations about socially structured variation from prior experience.

Crucially for this chapter, informativity provides a framework from which we can generate predictions about listener expectations under different simulations of the reference group. Namely, by manipulating the reference group from which socio-indexical structure is evaluated, we can identify the stability of socio-indexical factors and vowel categories under different hypothetical listener experiences. As such, using a large-scale dataset of American English, we can both replicate and expand Kleinschmidt (2019). The American English dataset used by Kleinschmidt (2019) is a limited representation of 'American English' composed of laboratory-elicited speech and generally homogenous beyond gender and regional dialect (see Clopper et al., 2005). As such, it is both a broad conceptualization of the baseline (i.e., American English) and yet a very narrow depiction of the variety overall. Theoretically, the use of a broad reference makes logical sense as the broader range of variation would require less frequent updating to listener representations and would represent conservative estimates of informativity. However, it is also necessary to examine other 'baseline' experiences to validate the method and generate testable hypotheses of the model. Furthermore, using 'American English' as a baseline, the range of speech should encompass a more representative and robust sample of the types of speech listeners are likely to encounter across talkers in order to fully validate the method and lend ecological validity to our hypotheses. I address this gap by examining different simulations of reference groups and drawing from a more diverse dataset. I will return to the details of quantifying previous experience in more detail in Section 2.3.

In addition to the question of previous experience, Kleinschmidt (2019) provides some initial predictions about how much socio-indexical structure can be identified across vowels. In Kleinschmidt's work, he describes a holistic dialectal structure which generalizes across vowels at two levels, which I refer to as *dialect-agnostic* and *vowel-agnostic*. Dialect-agnostic refers to the fact that informativity of the social factor, dialect, is a generalized aggregate property over all dialects and talkers. Analogously, vowel-agnostic refers to the fact that dialect informativity occurs across the vowel space where dialect is equally likely to be informative across all vowels at the broader contrast level (i.e., distinguished from, for example, stops). Namely, listeners build prior experience that suggests vowels vary across talkers as a function of their dialect background. Again, I refer to this as the *holistic* hypothesis, which describes generality over groups, talkers, and individual vowel categories and hypothesizes listener behavior across vowel categories as symmetrical.

Indeed, Kleinschmidt (2019) demonstrates that the degree of informativity of social groups, including dialect, is much greater for vowels compared to stops. Given these results, the holistic perspective predicts that all vowel categories result in similar adaptation and generalization behavior, whereby listeners are likely to treat any individual vowel category that deviates from prior experience equally by adapting flexibility to novel variation and generalizing to talkers of shared identity. Yet, as described in Chapter 2, there is evidence that vowels are not equally malleable in adaptation (e.g., Kataoka & Koo, 2017).

Conversely, Kleinschmidt (2019) also provides analytic results of how informativity emerges across more granular socio-indexical organizations from the combination of specific vowels and specific dialects. In the intermediate level, Kleinschmidt demonstrates dialect emerges as informative of the cue distributions of specific vowels that occur across regional shifts, such as /æ/; a *dialect-agnostic* but *vowel-specific* pattern. A further granular level is also described, which I refer to as *dialect-specific* and *vowel-specific*. In such cases, individual dialect areas (e.g., South) are informative of cue distributions of specific vowels (i.e., /eɪ/). However, such a division from a dialect-agnostic perspective is not developed theoretically within the ideal adapter framework, making the distinction, and ensuing predictions, of dialect-agnostic and dialect-specific organization unclear.

Furthermore, work describes the fact that vowels are highly talker-specific (e.g., Kleinschmidt, 2019; Samuel & Kraljic, 2009; Weatherholtz, 2015) as they contain a high degree of spectral information alongside factors of individual identity formation. In support, Kleinschmidt (2019) finds that talkers are more informative of cue distributions than dialect areas (and of gender). The distinction is suggested to indicate that when group is informative, perhaps regardless of magnitude, listeners will be more likely to generalize patterns to talkers of the same group. However, when group is not informative, then listeners are more likely to learn in talker-specific ways. There are many open questions about how these dichotomous classifications of socio-indexical structure (i.e., talkers vs. groups) affect listeners' representations and expectations.

Moreover, the informativity of socio-indexical structure requires that talkers within social groups are consistent and aligned with the broader group pattern. However, as detailed in Chapter 2, such alignment has been an open question in sociophonetics. Thus, even within a dialect-agnostic or dialect-specific organization, the degree to which talkers adhere to patterns of their dialect areas requires additional validation. I will largely structure the analysis in Section 4 around these different potential organizations of socio-indexical structure. In the next section, I will outline some expectations for patterns for each of the organizations based on current work in sociophonetics and more specific expectations about when talkers may diverge from their regional backgrounds.

2.2 Contextualizing Expectations in Production

Vowels represent an interesting nexus of socio-indexical structure where on the one hand dialect areas condition variation across vowel categories, demonstrate dialect-specific patterns, and illustrate a great deal of talker-specificity. This tension poses the necessity to validate the assumptions of ideal adapter models that groups are informative when they show high talker conformity to group norms. To guide the discussion, I will draw on traditional statistical terminology to facilitate interpretation. Specifically, I will use the term *factor* to refer to the higher order grouping variable (i.e., 'dialect') and the term *level* to refer to the components of a factor (i.e., 'South', 'West', etc.). Such terminology will extend to the other social variable of interest (individuals), with term *factor* referring to the aggregate group of talkers, and the term

level to refer to an individual talker. In the analyses in Section 4, the terminology reflects an analytic assumption of the separation between the factors of interest (i.e., dialect OR talkers). Each factor is regarded as explaining the variance of the joint cue distributions (F1xF2) as a composite of their respective individual levels, but there is not a hierarchical organization of talkers within dialects.

Subsequently, I will also refer to individual talkers within dialect areas, referring to a more nested structure to ask questions about the nature of individuals within their regional dialects. In such cases, I will use the term *nested* merely to distinguish from the flat structure associated with the dialect and talker factors; a visual illustration of the terms and the relationships is presented in Figure 4.2. I will refer to *vowels* broadly when referring to the aggregate of vowel categories (i.e., F1 and F2 across the vowel space), and individual vowel categories where relevant. The given predictions of socio-indexical structure in production will be the main focus in this section, I will return to the implications for Bayesian models of socio-indexical structure and perceptual learning in more detail in the overall discussion following the results (Section 5).

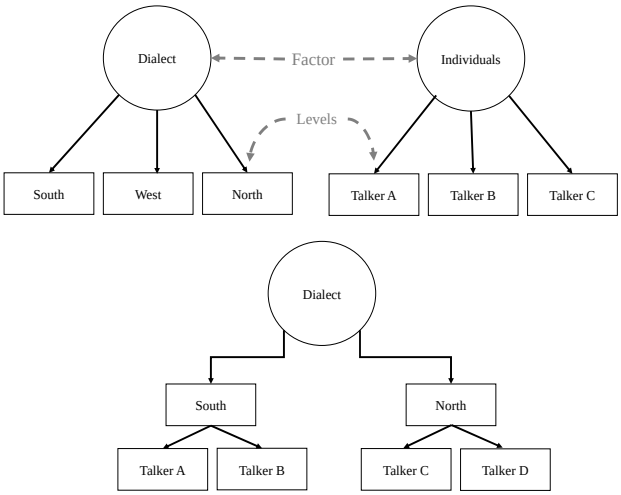

Figure 4.1 Socio-indexical organization based on more traditional terminology for clarity. The organization is meant as a description rather than a theoretical mental model.

In alignment with the dialect and vowel agnostic perspective outlined in Section 2.2, I hypothesize replicating Kleinschmidt's (2019) findings. Specifically, variability across vowel categories will be conditioned by the factor of Dialect, making the dialect factor informative of

the joint cue distributions of vowels broadly. Similarly, there is good reason to believe that the Talker factor is likely to condition variability to a greater degree than Dialect. These two components will ideally remain consistent across all vowel categories in comparison to the overall distribution of American English vowels. However, beyond that there are several other hypotheses we can make in terms of the granularity of expected contributions of individual dialect levels (e.g., 'South') and individual vowel categories (e.g., /æ/). Furthermore, assuming a more nested structure we should expect that individual talkers will diverge from their dialect areas in meaningful ways.

Pivoting from the dialect-agnostic and vowel-agnostic perspective, we now assess a more nuanced perspective of structure under a dialect-specific and/or vowel-specific framing. First, given the documented patterns of vowel shifts (see Chapter 2) I predict a *dialect-agnostic* and *vowel-specific* structure to emerge from the analyses in this chapter (Section 4), such that the Dialect factor conditions variability on an individual vowel basis. Given the fact that some vowel categories are more likely to be implicated across several shifts, we may likely see the generalization of dialects providing information to cue distributions for specific vowel categories. For example, /æ/ and /ɛ/ are categories that are more likely to vary across regional locales (Labov et al., 2006) making them robust examples of dialectally conditioned variability. Vowel categories that are likely to fit such a description are /æ/, /a/, /ɔ/, /ɛ/, and /ɪ/, where each vowel has been shown to vary by central tendency across regional dialects and is not limited to variation within a single regional variety. In such cases, we might predict listener behavior to be asymmetrical, such that categories like /æ/ will demonstrate more malleability in adaptation and greater generalization than categories like /i/ (see Section 5 and Chapter 6).

Next, we can hypothesize a more granular structure of *dialect-specific* and *vowel-specific* patterns that arise as informative. In such cases, there is limited generalization occurring over dialect areas or vowels and listeners generate more specific expectations about the cue distributions for individual vowel categories and dialect areas. For example, listeners may generate expectations that, in the West, the cue distributions of /æ/ are divergent from American English more broadly. As such, we could expect categories to emerge that are implicated only in a particular regional variety, or alternatively, when they are most divergent from the reference group (see Section 5 for more specific listener behavior predictions).

Both the *dialect-agnostic* and the *dialect-specific* hypotheses outlined thus far assume that dialects are not equally likely to condition cue distributions for all vowel categories in an informative way. When dialects are informative of cue distributions, either for the factor or at individual levels, I will refer to this overarchingly as *dialect-informative*. Consequently, the expectation would be that in the absence of dialectal conditioning, individual talkers' cue distributions are highly regular, and thus talker information is high. In cases where cue distributions are conditioned on talkers but not on dialect, I will refer to this pattern as *talker-informative.* While we expect talkers to always be generally more informative than the group, talker-informative specifically refers to cases where there is limited (or minimal) evidence of dialect conditioning for the same category.

The former predictions assume that when dialect (factor or levels) is informative of a vowel category's cue distribution it reflects greater homogeneity among talkers' cue distributions within their respective dialects. In other words, individuals will meaningfully group into regional dialects and demonstrate uniformity in cue distributions mirroring their community patterns. However, such an assumption is never directly evaluated and remains as an open question. The methods used by Kleinschmidt (2019) for aggregation of individuals to dialect areas occur as a flat structure (i.e., raw distribution of all tokens and talkers) making it difficult to validate such an assumption. Given the reality that individuals will diverge more or less from their group's norms (see e.g., Horvath & Sankoff, 1987), we can make some predictions about talkers as nested within regional groups.

As a vowel category undergoes change or gains social salience, there may be greater spread of talker norms within the community as individuals may have more awareness of variation and draw on phonetic resources in variable ways to index identity (Eckert, 2008, 2012; Erker, 2017; Guy & Hinskens, 2016; Trudgill, 1986). Given this, we might expect that vowel categories more saliently associated with regional vowel shifts are less likely to show robust dialect conditioning in the aggregate as a function of higher within-region variation. Thus, we might observe asymmetry across vowels such that some are dialectally conditioned (i.e., high regularity of talkers within regions) and categories that are talker-specific (i.e., higher idiosyncratic tendencies, more variability within regions). Such patterns may be evident when we

examine a more nested structure, examining individuals in relation to their dialect areas, rather than the flat structure provided by the Dialect and Talker factors and the respective levels.

Ongoing work of regional vowel shifts illustrates the variability that may arise across analyses. Recent work has demonstrated some regional vowel shifts are reversing, resulting in more heterogeneity between talkers. In the South, for example, talkers are not participating in the Southern Vowel Shift (SVS) to the same extent as evident by individuals within the region demonstrating variable participation and alignment with Standardized American English forms (Dodsworth & Kohn, 2012; Dodsworth, 2018; Kendall & Fridland, 2012). As a result, the within-region variation across talkers is greater for vowel categories predominately marked for the SVS, such as /eɪ/, /ɛ/, and /aɪ/. As such, we might hypothesize that /eɪ/, /ɛ/, and /aɪ/, are less likely to emerge as dialect-informative and demonstrate greater divergence of individuals from the reference group of the South. If /eɪ/, /ɛ/, and /aɪ/ are predominately associated with the SVS then we might expect that there is greater heterogeneity among individual talkers within that dialect area. And indeed, this could at least in part explain why Kleinschmidt (2019; see also Clopper et al., 2005) didn't observe a high degree of informativity of the dialect area for SVS vowels. The South is not the only region where such reversal has been observed. Recent work has demonstrated the North undergoing reversal of the NCS for /æ/ and /a/. Consequently, we might expect greater diversity among talkers within the Northern dialect region as well for these vowel categories. A prediction stemming from these patterns is that in regions where vowel shifts are undergoing reversal, there may be a decreased likelihood of seeing dialect-specific and vowel specific patterns that align with regional shifts. Instead, we may see evidence of greater talker-specificity for each of the vowel categories as indicated by the high informativity of talkers with respect to their dialect areas.

In addition, some vowel categories within a regional dialect are more likely to vary in their social meaning, and as such we might expect broadly more variability across talkers as they draw on different social resources during speech production. For example, /æ/ has been associated with variable social meanings in both production (Eckert, 2008; Podesva et al., 2015; Podesva, 2011) and listeners' perception of talker characteristics (D'Onofrio, 2015; Villarreal, 2018) across dialect areas. Consequently, there may be minimal dialect-specific informativity of /æ/ but high talker-specificity as talkers draw on the social meaning in variable ways. While the

dataset used in this chapter doesn't necessarily capture stylized speech, there may nonetheless be a broader range of variability from talkers with respect to their regional groups in their average behavior. Alternatively, categories that carry less social meaning or salience may be more likely to show higher within-dialect uniformity as a result of their unmarked status. For example, back vowel merger is not considered salient (Labov, 1994, Eckert & Labov, 2017) and thus may be more likely to show regularity across talkers within a given region. In such cases we might expect higher dialect-specific informativity and talkers within the region demonstrating minimal divergence from their groups in a nested model.

In the expectations outlined in this section, it is clear that there are any number of ways that social factors can be informative to vocalic cue distributions. Beyond the dialect-agnostic and vowel-agnostic framing of Kleinschmidt (2019), any degree of dialect or vowel specificity suggests asymmetries across vowel categories about which socio-indexical factors are informative of cue distributions. Given the fact that vowels will always be highly informative of talker identity, it's fruitful to consider the relational constituents of individuals and dialect areas. A fallout from this relational perspective is a division between group-informative and talker-informative patterns, where group-informative patterns demonstrate consistent regularity within dialect areas and talkers show minimal divergence from their regions. Similarly, talker-informative patterns would show regular null effects of dialect informativity and greater between-talker variability within regions. The source of group-informative and talker-informative vowels may be correlated with social salience of the category where more salient categories show less regularity across talkers than others. Such a distinction is important because this potential asymmetry across vowels makes them a unique test case for understanding how listeners' a priori knowledge of socio-indexical structure may produce different adaptation and generalization behavior. I will focus primarily on the cases of dialect and vowel-specific descriptions of socio-indexical structure in the analyses in Section 4 and return to discuss implications for listener behaviors in the discussion.

2.3 Quantifying Previous Experience

In this chapter, previous experience is simulated in three ways to assess the degree to which vocalic variability shifts in dialect and talker informativity under different 'baseline'

experiences. The analyses are not meant to represent an exhaustive account of different listener experiences and do not account for experience with second language English speakers, ethnic varieties, or other varieties of English outside of the U.S. However, conducting simulations with a corpus of natural speech allows us to generate testable hypotheses and begin validating the assumptions of ideal adapter models under more ecologically valid representations than lab speech (see Chapter 6 as one example). In the first analysis, I simulate a robust experience with American English, to validate informativity of socio-indexical factors (Analysis 1, Section 4.1). In the second analysis, I will examine a smaller subset of data comprised of regions that represent the regional vowel shifts more specifically (Analysis 2; Section 4.2). In the third analysis, I shift perspectives and assume that listeners' baseline means of evaluating talker structure is in reference to their variety and simulate this by examining divergence from a single region as a reference group (Analysis 3, Section 4.3). In the following paragraphs, I will describe the motivation for each of these choices in more detail and provide some general expectations.

In the first analysis, a robust experience with American English is simulated using the entirety of the datasets outlined in Chapter 3. This larger dataset represents an exposure to American English that is representative of speech listeners are more likely to encounter where it may be imbalanced across regions, across speakers, and a diversity of speech styles. This analysis provides insight into a more dynamic representation of speech than what is typically experienced in a laboratory setting and provides a more accurate reflection of the diversity listeners contend with. Such a test case is essential to validating both the methodological utility of KL divergence and for validating the computational model's predictions more broadly. From this perspective, the baseline experience of 'American English' provides the most conservative estimate of socio-indexical informativity across different analytic organizations of social structure and under the more representative noisy sampling distributions.

In analysis two, I analyze a subset of the data that represents a more constrained model of American English that is represented by regions that participate in the most studied and described vowel shifts (North, South, West). This functions in part to validate KL divergence with a more controlled (i.e., less noisy) sample of conversational speech where there are clearer expectations based on extensive work into these regions. This baseline experience represents a sort of middle ground, where the reference encoding represents the margins of the different

regional vowel systems. Primarily, this treatment serves to understand whether the patterns observed in Analysis 1 can be attributed to the overall diversity of the corpora. We also might expect that informativity is likely to be emergent when examining regional patterns that are expected to be most divergent, and indeed the availability of structure to listeners may mirror similar expectations.

Finally, I take a narrower view of experience where listeners' primary exposure is of a single dialect area (e.g., West) and ask how informativity of vowel categories may hinge on listeners' prior experience with their own dialect area. Some work has suggested that listeners make perceptual evaluations in relation to their own dialect areas (e.g., Fridland 2008) and have demonstrated dialect categorization is influenced by mobility in early linguistic experience (Clopper & Pisoni, 2004b). Thus, it's not unreasonable to assume that listeners may be identifying socio-indexical structure in relation to their own dialect areas, rather than more broadly construed 'American English' or a wide-ranging set of experiences with regional vowel systems. This expands on the tests in the prior section by probing a different model of the 'baseline' data and diverges from the Kleinschmidt (2016, 2019) by hypothesizing that potential structure is evaluated in relation to a baseline community model. This simulation also provides an initial depiction of how dialect areas diverge from one another in meaningful ways, which has methodological implications for a range of sociophonetic questions beyond those outlined here. Overall, these three simulations of priors usefully inform predictions about what listeners may reasonably identify to form a priori assumptions about dialect and talker variation.

3 Methods

3.1 KL Divergence: Theoretical Details

Throughout this chapter I will use KL divergence as a measure of socio-indexical structure, following work by Kleinschmidt (2019). At a more general level, KL divergence is an asymmetrical measure of divergence (or distance) between two probability distributions, as denoted by the *D* in D(*Q*||*P*). KL divergence is a well-attested metric used across disciplines interpreted variably as measure of relative information, of uncertainty, or of surprisal and salience (Commenges, 2015). In information theory, relative entropy describes the relative information loss associated with encoding a true probability distribution (*P*) with an estimated

prior distribution (*Q*; Kullback & Leibler, 1951). Or, in other words, how much uncertainty about the cue distributions is reduced by knowing *P* compared to *Q*. In this chapter, I follow Kleinschmidt (2019) in interpreting KL divergence values as denoting information gained by socio-indexical factors (*P*) in relation to an estimated reference distribution *Q*, which varies across analyses. Though, I will also use 'divergence' as a simplified atheoretical term to describe the patterns in the data when relevant. To summarize, high divergence denotes a loss of information when using a distribution *Q* (e.g., American English) to estimate a true distribution *P* (e.g., a talker's true distribution) demonstrating there is more information gained by knowing the true distribution *P* (e.g., talker).

KL divergence is not meant to represent a cognitively real measure of what listeners track or represent, but rather acts as a means to quantify potential socio-indexical structure across cue distributions and identify candidate vowels listeners have prior experience with as being socially conditioned. In this chapter, Dialect and Talker are the primary socio-indexical factors of interest, however, I also replicate Kleinschmidt (2016, 2019) by calculating Gender and Dialect+Gender in Analysis 1 and 2 (Section 4.1- 4.2). I expand the use of KL divergence in two ways: 1) evaluating how much information is gained by talkers' cue distributions relative to their dialect area cue distributions, and 2) evaluating how much information is gained by socio-indexical factors when comparing a single dialect area (*Q* = West) to other dialects (e.g., *P* = South; see Richter et al., 2016 for similar methods) and Talkers (e.g., *P* = talker01). The first addition tests the internal consistency of individuals within groups—a critical piece of ideal adapter models and the potential of informativity for a particular factor. In addition, it provides an assessment of how much information would be lost if listeners were to a priori use a dialect area to estimate individual talkers within those groups. The second addition allows us to identify how much structure is evident when listeners use their own community as an estimate for talkers and dialects (Section 4.3, Analysis 3).

3.2 KL Divergence: Technical Details

As described above, KL divergence is a non-symmetric measure of the difference between two probability distributions $P(x)$ and $Q(x)$, where x represents a random variable, here as the multivariate distributions of F1 and F2 conditioned on individual vowel categories.

Following Kleinschmidt (2019), $P(x)$ and $Q(x)$ are assumed to be normal multivariate cue distributions conditioned on vowel category parameterized by the mean and covariance; $P(x)$ is also conditioned on socio-indexical factor (e.g., dialect area, talker, etc.). The KL divergence values will be reported in *bits* for ease of interpretation; a value of 0 indicates identical distributions and greater values indicate more information gained by the socio-indexical factor. The analyses here are not a direct borrowing of the computational model itself and may be best understood as a method of describing different socio-indexical scopes and baseline conditions. KL-Divergence was computed by adapting the phondisttools package (Kleinschmidt, 2019) for the use cases below. For a full overview of the technical and mathematical details of KL divergence, see Kleinschmidt (2019).

In the analyses throughout Section 4, $Q(x)$ will vary in terms of underlying data (see Section 2.3) but will similarly be represented by a normal multivariate cue distribution conditioned on vowel category and parameterized by the mean and covariance. For illustrative purposes, Figure 4.2 provides a demonstration of the distinction between the marginal distribution *Q* and the true distribution *P.* The highlighted red ellipses represent the dialect area 'South' as $P(_{\text{F1xF2 | vowel, South}})$ and the grey ellipses represent the marginal distribution $Q(_{\text{F1xF2 | vowel, all}})$. KL Divergence will estimate how much the marginal (Q, grey) distribution diverges from the 'true' distribution, South (P, red) for each vowel category.

In each of the analyses below, I use different reference distributions (*Q*), comprised of subsets of the dataset in Chapter 3, simulating different listener priors under which socio-indexical structure is evaluated. In Analysis 1, $Q(x)$ will be a marginal distribution of all data from Chapter 3, conditioned on vowel category, but not socio-indexical group: $Q(_{\text{F1xF2 | vowel, all}})$ this distribution will be denoted as Q_M. In Analysis 2, $Q(x)$ will be the marginal distribution of a single dataset, the Switchboard corpus, and three dialect areas (North, South, West) conditioned on vowel but not socio-indexical group: $Q(_{\text{F1xF2 | vowel, Switchboard[N,S,W]}})$ this distribution will be denoted as Q_S. Analysis 1 and 2 will also have sub-analyses (1b: Section 4.1.5 and 2b: Section 4.2.5) where $Q(x)$ will be represented by individual dialect areas, $Q(_{\text{F1xF2 | vowel, dialect-}i})$, and $P(x)$ will represent individual talkers (Sections 4.1.5 and 4.2.5); for Analysis 1b these groups will be denoted as Q_{Md} and Analysis 2b will be denoted as Q_{Sd}. In Analysis 3, $Q(x)$ will be a single

dialect region, the West, conditioned on vowel: $Q_{(F1xF2 \mid vowel, West)}$ this distribution will be denoted as Q_W.

The distributions of $P(x)$ will be conditioned on vowel category and socio-indexical factor levels, with the following conventions P_{Di} as dialect levels, and P_{Ti} as individual talkers. In the analyses to follow, KL divergence is calculated for each level of the socio-indexical factor of interest conditioned on each vowel category (e.g., $D(Q_M||P_{di})$, $D(Q_M||P_{ti})$). Reported KL divergence for a factor (e.g., Dialect or Talker) represents an average of KL divergence values over the respective levels. This method follows Kleinschmidt (2019) and is used for convenience and parity with the methods therein. To evaluate KL divergence for a given socio-indexical factor or level, I compare the true values to a comparable group composed of random shuffling of talkers to social group labels and tokens (grouped by vowel category) to talker labels. For example, individual talkers were randomly assigned to two groups to mirror the conditioning of gender. Similarly, talkers were randomly assigned to groups with the labels of dialect areas mirroring the total talker counts in the true group labels. This is merely meant to provide a baseline random sample for comparison, which should provide relative strength for the informativity of the true groups. While a more conservative estimation using bootstrapping confidence intervals would be preferred, the overall size of the dataset made bootstrapping computationally costly and is left for future work. As such, KL-Divergence is treated as descriptive statistics for various analytic groups and is not evaluated for statistical significance.

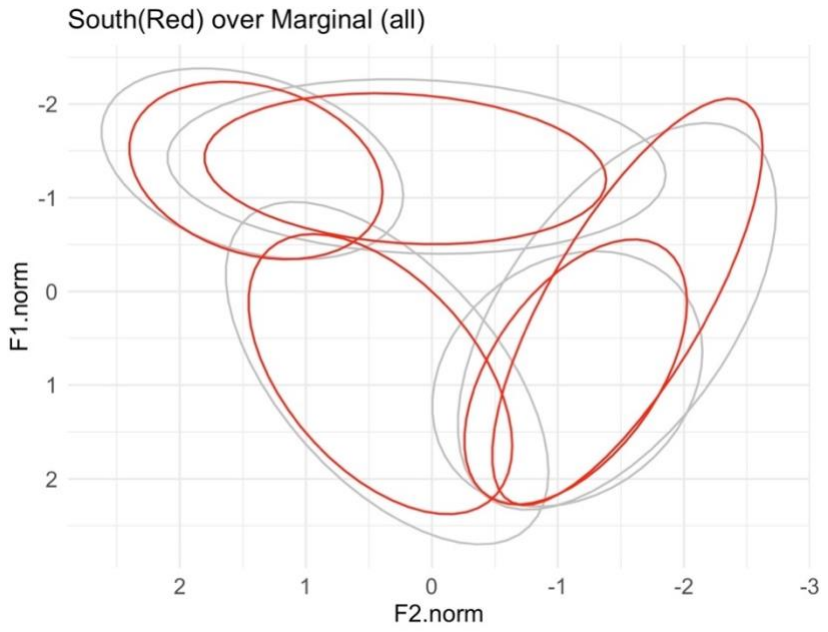

Figure 4.2: Example of socio-indexical group distributions (red) over the marginal distributions (grey) for American English broadly.

4 Analyses

Given the descriptions above, the analyses will be organized around the three 'baseline' experiences: American English (all data; Analysis 1), Shift Specific Regions (subset data; Analysis 2), and Single Community Baseline (Analysis 3). For each of these different 'baseline' analyses, I will present the results in the following structure: first an overview of the high-level patterns, then dialect-agnostic and vowel-specific patterns (e.g., dialect is informative of /eɪ/), followed by dialect-specific and vowel-specific patterns (e.g., the West is informative of /æ/). In Analyses 1 and 2 I also present findings from what I am describing as a nested model and ask how much information would be gained by attending to individual talkers if listeners treated each talker according to their dialect area. If listeners approached the task knowing the talkers' regional dialect, how much additional information would be gained from talker identity. This sub-analysis primarily serves to evaluate how much homogeneity can be assumed for a given regional background and vowel category, a core axiom of the model. Following these results, I will turn to a broader discussion describing implications for perceptual learning.

4.1 Analysis 1: American English Baseline (All Data)

Turning to the first analysis, this section assumes maximal exposure to American English, encompassing seven different dialect areas as well as a range of types of speech and individuals within each region. The results of this baseline model are hypothesized to align with the results of Kleinschmidt (2019) due to the similarities in the assumed exposure to a broad range of American English. Following Kleinschmidt (2019), I expect to see that socio-indexical groups are informative of vowel categories, with more specific groups showing higher socio-indexical informativity, with a relative ranking from least to most informative as follows: gender < dialect < combination of dialect + gender < individuals. Based on the discussion in Section 2, I further hypothesize that dialect-agnostic and vowel-specific patterns are likely to emerge whereby the dialect factor is more informative of vowel categories associated with *several* regional vowel shifts, as for example /æ/.

4.1.1 Data & Method

In this analysis, the data are drawn from all corpora represented in Chapter 3: Data & Methods. In the following analyses of KL Divergence, the following factors and levels are assumed: marginal distribution, $Q_M(_{\text{F1xF2 | vowel, all}})$, as described in Section 3.2; Dialect factor with 7 levels, $D(Q_M||P_{di})$, gender with 2 levels, $D(Q_M||P_{gi})$, and Dialect+Gender with 14 levels (7 dialect levels x 2 gender levels) $D(Q_M||P_{dgi})$, and Talker with 957 levels (i.e., individuals), $D(Q_M||P_{Ti})$. Each social grouping along with token counts and number of speakers are provided in Table 4.1. The marginal distribution in the analyses to follow encompasses all data available, conditioned on vowel category; Table 4.2 provides the total number of tokens for each categories' marginal distribution. Values of KL Divergence for the socio-indexical factors are calculated for each level of the factor and then averaged over levels to provide an overall informativity for each socio-indexical factor. For example, when calculating KL Divergence for 'Dialect' each dialect level (e.g., South, West, etc.) will have a value of informativity for each category (e.g., /æ/) which is then averaged to give an overall informativity of the dialectal factor for each vowel category (following Kleinschmidt 2019). In Section 4.1.2, I will report the factor averages for each vowel category (e.g., dialect informativity for /æ/); in Section 4.1.4 I report the

values for each level of the dialect factor for each vowel category. For each socio-indexical group a random assignment of talkers to groups, or tokens to talkers, is represented for an evaluation of the 'null'. Specifically, talkers are randomly assigned to 7 groups (equal to the number of dialects), and again to 2 groups (gender), and 14 groups (dialect X gender).

Table 4.1 Total unique speaker counts for each dialect area and gender across all data presented in Chapter 3.

Dialect Area	**Men (N)**	**Women (N)**	**Dialect (N)**
Midatlantic	5	9	14
Midland	153	151	304
North	37	21	58
Northeast	108	108	216
NYC	11	8	19
South	109	103	212
West	58	59	117
Total N	**481**	**459**	**940**

Table 4.2 Total token counts per vowel category for the marginal distribution (Q_M).

Vowel	**N**
a	36891
æ	41228
ʌ	46536
ɔ	52057
aɪ	62260
ɛ	65666
eɪ	51755
i	58304
ɪ	40186
o	58422
u	32904
ʊ	7154
Total	**556946**

4.1.2 Higher-Order Factors

Figure 4.3 illustrates the mean KL divergence of different socio-indexical factors from the marginal data (all tokens across all datasets). Higher values of KL divergence indicate greater informativity of cue distributions, and a lower KL divergence value indicates greater similarity between the two distributions and thus less information gained by knowing the 'true' distribution (dialect, talker) compared to the marginal (American English). Again, each socio-indexical factor (indicated by color on the figure) represents the factor *average* of KL divergence over the respective levels. Empty circles represent a random assignment of talkers to dialects and tokens to talkers, and then similarly averaged by the randomized factor label.

Based on the findings of Kleinschmidt (2016, 2019) we expect that more granular factors will be more informative of cue distributions, and indeed such findings are replicated here, such that KL divergence is highest when the factor is most granular (talkers) and lowest when the factor is broadest (gender). On average informativity varies as a function of socio-indexical group, where individual talkers provide the most information (mean KL = 0.99 bits), followed by a combination of dialect and gender (mean KL = 0.14 bits), then dialect (mean KL = 0.10 bits), followed by gender as the least informative (mean KL = 0.01). Comparing these results to a random assignment of talkers to groups (Random-Dialect, Random-Talker, etc.) generally demonstrates that the real groups provide more information gain over the random groupings, though for gender the true value is already so minimal it likely is not a true effect[4] (Random-Gender mean KL = 0.001). Additionally, the ranking of least to most informative socio-indexical groups remain the same for random groupings of talkers (and tokens to talkers), suggesting that more informativity of more specific groupings may be a product of the measure. Finally, talker identity is still more informative than a random shuffling of tokens within phoneme to 'talker identity' grouping levels, confirming that talker identity is a reliable effect.

[4] It is worth noting that since the data are Lobanov normalized, the gross difference between men and women has been normalized out—we would expect to see that gender is more informative of vocalic patterns in raw Hz (see Kleinschmidt 2016, 2019 for evidence).

In the following sections I will turn to examine vowel-specific patterns. Given the high-level results largely replicate Kleinschmidt (2019), and the primary focus of this dissertation is on dialect groups and individual talkers, for the remainder of the chapter I will focus only on these two social factors. In addition, given the overall ranking of Talker as higher in informativity compared to Dialect, I will focus on the relative ranking of vowel categories within each socio-indexical factor in the following vowel-specific analyses. This merely aims to aid in interpretation of the findings and is not meant to indicate that the highest KL divergence value for dialect will somehow 'win' in perceptual processes but identifies any asymmetries between the categories.

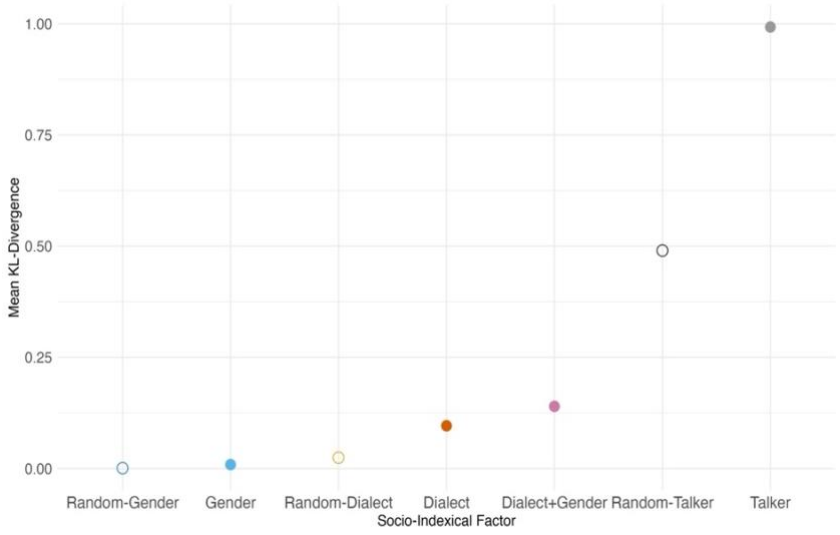

Figure 4.3 Mean KL divergence for each socio-indexical factor (filled circles), including randomly assigned talkers (& tokens) to comparable sized groupings (unfilled circles). Averaged over respective levels and vowel categories.

Table 4.3 Mean KL divergence for each socio-indexical factor and randomized groups.

Factor	Mean KL
Talker	0.99
Random-Talker	0.49
Dialect+Gender	0.14
Dialect	0.10
Random-Dialect	0.02
Gender	0.01
Random-Gender	0.00

4.1.3 Vowel-specific: Dialect-Agnostic & Talkers

Figure 4.4 depicts the average KL divergence value for Talker (grey filled) and Dialect (orange filled) factors, alongside the random factors (unfilled circles) by vowel category. Again, higher values indicate greater conditioning of variability for the individual vowel categories' cue distributions and more information gained. First, as expected, Talker remains an order of magnitude higher in informativity than Dialect for each of the vowel categories. Given the discussion above (Section 2; see also Chapter 2) we might expect that Dialect and Talker will be asymmetric in informativity across vowel categories, such that vowel categories implicated *across* regional shifts (e.g., /æ/) are more likely to emerge as ranked high in Dialect information than those that are not associated with regional shifts (e.g., /ʊ/). Similarly, categories that are not implicated across regional shifts may be more likely to emerge as ranked high in Talker information, as Dialect, broadly, is not expected to condition variability.

Table 4.4 provides mean KL divergence values for Dialect and Talker factors, respectively, rank ordered. In Figure 4.4 we can see there are a few vowel categories which are near equal in KL divergence for Dialect and a random assignment of talkers to equal groups, including /ʌ/, /ɪ/, and /u/, similarly they are ranked lowest in Dialect information in Table 4.4. This pattern suggests that dialect groups do not provide more information than the marginal in estimating cue distributions for all vowel categories. On the other hand, KL divergence of talker

information is always greater than a random assignment of items to talkers. Table 4.4 and Figure 4.4 illustrate evidence of the high-level prediction that Dialect and Talker informativity will be inversely related, such that categories ranking highest in Dialect information do not emerge as highly ranked in Talker information, and vice versa. Specifically, the vowels /ɔ/, /aɪ/, and /eɪ/ are ranked as highest for Dialect, with /aɪ/ and /ɔ/ falling near the bottom of the rank ordering and /eɪ/ somewhere in the middle for Talker. Similarly, /ʊ/, /ʌ/, and /ɪ/ are ranked highest for Talker and are ranked lowest in Dialect informativity.

However, the categories highest ranked in Dialect do not appear to be strongly correlated with expectations of regional vowel shifts, with the exception of /ɔ/ which may be largely indicative of low back vowel merger patterns across regions. Nonetheless, the relationship between Dialect information and Talker information suggests that for some categories there is stronger conditioning on Dialect and Talker may not provide a gain over and above Dialect. Of course, this is only weakly evident from the results here, but I will return to this point again in Section 4.1.5 to further investigate the claim quantitatively. The fact that Dialect did not emerge as informative for categories associated with vowel shifts may, in part, be driven by the fact that the factor is too broad and informativity of such socially meaningful vowels is more likely to occur in a dialect-specific manner rather than the dialect-agnostic perspective provided here.

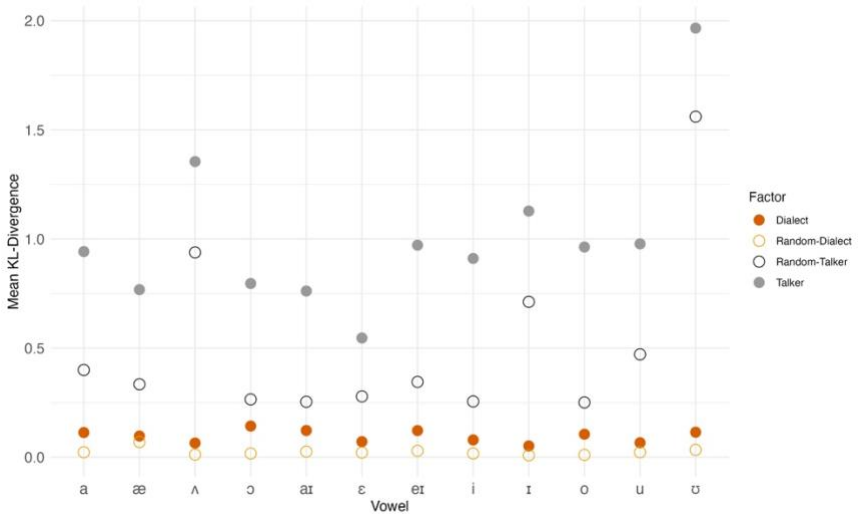

Figure 4.4: Mean KL divergence for the socio-indexical factors of Talker and Dialect (filled circles), including randomly assigned talkers (& tokens to talkers) to comparable sized groupings (unfilled circles)

over marginal distribution. Averaged over respective levels and separated by vowel category.

Table 4.4: Mean KL divergence for the socio-indexical factors of Talker and Dialect over marginal (all) distributions, rank ordered respectively.

Dialect		**Talker**	
Vowel	**Mean KL**	**Vowel**	**Mean KL**
ɔ	0.14	ʊ	1.97
aɪ	0.12	ʌ	1.35
eɪ	0.12	ɪ	1.13
a	0.11	u	0.98
o	0.11	eɪ	0.97
ʊ	0.11	o	0.96
æ	0.10	a	0.94
i	0.08	i	0.91
ʌ	0.07	ɔ	0.80
ɛ	0.07	æ	0.77
u	0.07	aɪ	0.76
ɪ	0.05	ɛ	0.55
Mean	**0.10**	**Mean**	**1.01**

4.1.4 Vowel-Specific: Dialect-Specific

We expect to see greater informativity of vowel categories associated with shifts within regions, regardless of whether they were informative of 'dialect' in the previous analyses.

Figure 4.5 present the results of KL divergence across vowel categories and the seven dialect areas. Before moving to the specific vowel categories of interest, it's first worth noting that on average each dialect area ranges in informativity across vowel categories. For example, the Midland (mean = 0.02 bits) and West (mean = 0.05 bits) dialect areas demonstrate relatively low informativity in relation to the marginal distributions across all vowel categories. In contrast, the Northeast (mean = 0.21 bits) and the Midatlantic (mean = 0.15 bits) demonstrate greater informativity across many of the vowel categories, while the North (mean = 0.09 bits) and South (mean = 0.08 bits) fall somewhere in between. Within the regions demonstrating greater informativity on average, the lowest individual category values are often higher than even the most informative categories in the Midland and the West areas. For example, the highest ranked

category for the Midwest is less than 0.10, while almost all categories in the South are greater than 0.10. Overall, this confirms that some dialect areas' cue distributions are more divergent from the marginal 'American English' values. This provides quantitative support of a more standardized or 'general' American English variety that aligns more closely with speech in the Midland and West and aligns with listener evaluations of where the most standardized variety is spoken (Preston, 2011).

If we further examine within dialect rankings for specific vowel categories, we see some trends in line with expectations of regional vowel shifts, illustrated in

Figure 4.5 and Table 4.5. In particular, in the West we see higher informativity of /a/ and /æ/ (0.07 bits), in-line with shifts in the low back vowels associated with the LMBS. In the South the vowel categories with highest informativity are /aɪ/ (0.24 bits) and /ɔ/ (0.10 bits), in alignment with the SVS and a more conservative /ɔ/ position (Fridland & Kendall, 2015; Thomas, 2001; Labov et al., 2006). Further, such an observation aligns with social perceptions of Southern /aɪ/ (Albritten, 2011; Plichta & Preston, 2005) and evaluation of Southern talkers' /ɔ/ as accented (Gunter et al., 2020). In the North the categories higher in informativity are /aɪ/ (0.25 bits), and /o/ (0.18 bits). While these categories are not related to the NCS patterns, they still demonstrate regional patterns in the North. Northern /aɪ/ demonstrates more robust raising in the North compared to some other regions (Labov et al., 2006). Likewise, /o/ in the North may be driven by the fact that /o/ fronting is prevalent across some varieties of American English (Labov et al., 2006) while the North maintains a more backed variant of /o/ (Labov 1994; Labov et al., 2006). NYC also retains a more conservative backed /o/ (Labov et al., 2006) and similarly is higher in informativity of /o/ distributions (0.23 bits) compared to other vowel categories. However, there are some results that are inconsistent with expectations of regional patterns. For example, the West is more informative of /eɪ/ (0.14 bits), despite /eɪ/ being relatively stable in the region (e.g., D'Onofrio et al. 2016), and the Midatlantic shows greater informativity of /ʊ/ (0.39 bits) and /ʌ/ (0.21 bits). The informativity of the Midatlantic vowels may be indicative of smaller sample sizes, both in terms of unique talkers and overall number of tokens from talkers on average compared to other regional varieties.

However, the explanation for informativity of the West for /eɪ/ is less clear but aligns with the more generalized dialect factor (averaged over dialect levels) showing informativity of

/eɪ/ distributions (see Section 4.1.3). The dialect-informative (i.e., both dialect-agnostic and dialect-specific) and vowel-specific patterns observed in these two sections overall demonstrates less than perfect mapping of dialect informativity of cue distributions for vowel categories associated with regional vowel shifts. The fact that /eɪ/ distributions seem to be conditioned on regional identity suggests that informativity may capture categories where talkers within regions are more regular in their cue distributions which may or may not align with the canonical expectations of regional vowel shifts. One explanation for the absence of the expected regional patterns in this analysis may be the fact that regions are internally variable, thus any degree of vowel-specific patterns we expect for regions may not emerge due to the higher degree of cross-talker heterogeneity across more socially variable vowel categories. In the next section I specifically aim to tackle this question, asking how much individuals' distributional patterns are divergent from their regions, thus making talkers emerge as more informative than regional identity for categories with regional affiliations.

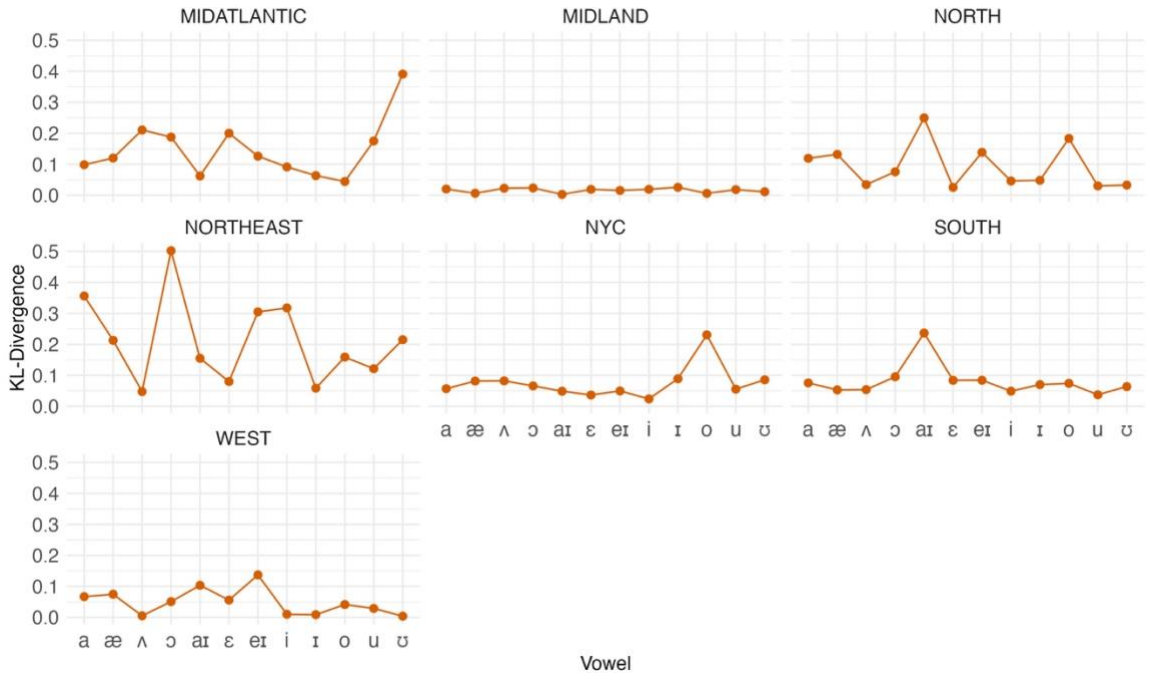

Figure 4.5 KL divergence values each level of the Dialect factor (i.e., individual dialect areas) by vowel category.

Table 4.5 KL divergence values each level of the Dialect factor (i.e., individual dialect areas) by vowel category, rank ordered within regions.

Midatlantic		**Midland**		**North**		**Northeast**	
Vowel	**KL**	**Vowel**	**KL**	**Vowel**	**KL**	**Vowel**	**KL**
ʊ	0.39	ɪ	0.03	aɪ	0.25	ɔ	0.50
ʌ	0.21	a	0.02	o	0.18	a	0.36
ɛ	0.20	ʌ	0.02	eɪ	0.14	i	0.32
ɔ	0.19	ɔ	0.02	æ	0.13	eɪ	0.30
u	0.18	ɛ	0.02	a	0.12	æ	0.21
eɪ	0.13	eɪ	0.02	ɔ	0.08	ʊ	0.21
æ	0.12	i	0.02	i	0.05	o	0.16
a	0.10	u	0.02	ɪ	0.05	aɪ	0.15
i	0.09	æ	0.01	ʌ	0.03	u	0.12
aɪ	0.06	o	0.01	ɛ	0.03	ɛ	0.08
ɪ	0.06	ʊ	0.01	u	0.03	ɪ	0.06
o	0.04	aɪ	0.00	ʊ	0.03	ʌ	0.05
Mean	**0.15**	**Mean**	**0.02**	**Mean**	**0.09**	**Mean**	**0.21**

NYC		**South**		**West**	
Vowel	**KL**	**Vowel**	**KL**	**Vowel**	**KL**
o	0.23	aɪ	0.24	eɪ	0.14
ɪ	0.09	ɔ	0.10	aɪ	0.10
ʊ	0.09	a	0.08	a	0.07
æ	0.08	ɛ	0.08	æ	0.07
ʌ	0.08	eɪ	0.08	ɛ	0.06
ɔ	0.07	ɪ	0.07	ɔ	0.05
a	0.06	o	0.07	o	0.04
u	0.06	ʊ	0.06	u	0.03
aɪ	0.05	æ	0.05	ʌ	0.01
eɪ	0.05	ʌ	0.05	i	0.01
ɛ	0.04	i	0.05	ɪ	0.01
i	0.02	u	0.04	ʊ	0.00
Mean	**0.08**	**Mean**	**0.08**	**Mean**	**0.05**

4.1.5 Analysis 1b: Talkers Within Dialects (Nested)

At this point, I turn to consider whether individuals' distributions align (or not) with the distributions of their groups. In this section the true distribution is an individual talker (P_{ti}), and the marginal distribution is the dialect area to which they belong, $Q_{SDi}(\text{F1xF2} \mid \text{vowel,di})$. For example, KL divergence is calculated for a single talker, Talker001, from a single dialect area, South, for each vowel category. As such, this section represents a new analysis using the same data, described in Section 4.1.1, as the 'baseline' is now relative to the individual talker's regional affiliation. The rationale for this analysis lies in the fact that current theoretical models assume some degree of homogeneity among talkers within dialect areas, thus this is a direct test of the extent of between-talker variation in dialect areas and represents an approximation to a nested structure.

If the homogeneity assumption outlined by Kleinschmidt (2019) is accurate, we should expect to see that for vowel categories ranked higher in dialect informativity in Section 4.1.3 (e.g., /eɪ/), the factor of talker should rank low, showing greater similarity to their dialect areas and little information gained. Similarly, if categories ranked high in talker informativity in Section 4.1.3 (e.g., /ʊ/) are indicative of high between-talker variability and absence of dialect conditioning then talkers should have higher KL divergence in those categories in relation to their dialect areas as well. Results displayed in Table 4.6 and Figure 4.6 below, provide some evidence for the assumptions outlined in Kleinschmidt (2019). Table 4.6 shows that on average, talkers diverge from their dialect areas for the same vowel categories in Section 4.1.3 where Talker information was high. In particular, /ʊ/ ranks second highest in talker divergence (mean KL = 0.64 bits), whereas /eɪ/ shows lower talker divergence (mean KL = 0.55 bits). In other words, in categories where dialect informativity is higher (demonstrated in Section 4.1.3-4.1.4), talker divergence from their dialect areas is ranked lower for those categories. Further, when using dialect area to estimate talkers' cue distributions, we see that the average KL divergence values are lower across the board, ranging from 0.40 to 0.67, compared to the use of the marginal for estimation (0.55-1.97). These patterns broadly contribute validity to the assumption that talkers are more aligned with their dialect areas than the marginal distribution, and for some vowel categories there appears to be higher dialect informativity and greater regularity among talkers within their dialect areas.

Table 4.6 Mean KL divergence for the Talker factor from their dialect areas' distributions, rank ordered. Averaged over individual talkers across regional backgrounds.

Vowel	Mean KL
i	0.67
ʊ	0.64
a	0.62
u	0.56
aɪ	0.55
eɪ	0.55
o	0.51
æ	0.51
ɔ	0.51
ʌ	0.43
ɛ	0.40
ɪ	0.40
Mean	**0.53**

4.1.6 Talker-Specific Patterns by Dialect

However, there may be dialect- and vowel-specific combinations where we might expect that talkers are more likely to diverge from their dialect areas. Such a distinction may further elucidate the patterns observed in Section 4.1.2-4.1.4 where vowel categories expected to be highly informative of dialect-specific patterns (e.g., Southern /ɛ/) may not demonstrate high informativity because talkers within a given region are highly variable. Such variability is hypothesized to affect predominately categories associated with high salience and/or regional shifts. For the South, the front tense-lax vowels are likely to demonstrate wider range of variability due to different degrees of participation in the SVS. For the North, /æ/ and /a/ may show higher variability as a function of salience and reversal of the NCS (D'Onofrio & Benheim, 2020; Driscoll & Lape, 2015; King, 2021; Nesbitt, 2021; Wagner et al., 2016). In the West, we may see greater variability in /æ/ given the variable and complex social meanings associated with the category (e.g., Podesva, 2011; D'Onofrio, 2016) and ongoing change in the region (e.g., Becker 2019). The average talker factor may not capture this in the overall ranking of vowel categories due to the dialect-specific and vowel-specific expectations. It may not be the case that

/æ/ demonstrates overall information loss across talkers from several regional backgrounds, but it may show that talkers diverge in specific regions where /æ/ has gained more prominent social salience or is used more for sociolinguistic style (e.g., the LBMS).

Turning to individual dialect regions, as illustrated in Figure 4.6 and Table 4.7, we see some patterns but not overwhelming evidence for this prediction. Aside from the categories that demonstrate highest Talker informativity in Section 4.1.3 (e.g., /ʊ/), some vowel categories show greater divergence for certain regions that may potentially be driven by variable participation in regional shifts. First, the South demonstrates expected patterns such that vowels related to the Southern Vowel Shift show greater divergence, including /aɪ/ (KL = 0.60 bits), /eɪ/ (KL = 0.57 bits), and /i/ (KL = 0.58 bits). As described above, this pattern could be due to the Southern Vowel Shift reversing among some talkers in the South (Dodsworth & Kohn 2012) and the fact that shifts in /i/ are restricted in geographical dispersion compared to other vowels implicated in the SVS (Labov et al., 2006; Fridland, 2012). While the South was informative of /aɪ/ variability in Section 4.1.4 above, it appears that there is still a high degree of inter-talker variability.

Additionally, /a/ appears as more informative of Talkers in the North (KL = 0.63 bits), which may be driven by variable participation in the NCS, stylistic variation (Eckert, 2000; Van Hofwegen, 2013), or gendered variation (Eckert, 2000). However, other dialect areas also show greater Talker informativity in /a/, including the Midland (KL = 0.71 bits) and West (KL = 0.59 bits), which may be explained in part by back vowel merger gaining prevalence (Labov et al., 2006). Comparably to /a/, /u/ also shows greater Talker informativity in the Midwest (KL = 0.62 bits) and West (KL = 0.55 bits), potentially related to advancement of /u/ fronting in those regions leading to greater between-talker diversity (Labov et al., 2006). However, overall, the categories most indicative of talker divergence are the ones in which talkers diverge from the marginal (e.g., /ʊ/) and predominately other categories that were not ranked highly in Dialect levels informativity in Section 4.1.3. Such a pattern suggests that when Dialect is informative, talkers within their dialect areas also demonstrate greater alignment with the dialect patterns confirming some of the assumptions outlined by Kleinschmidt (2016, 2019).

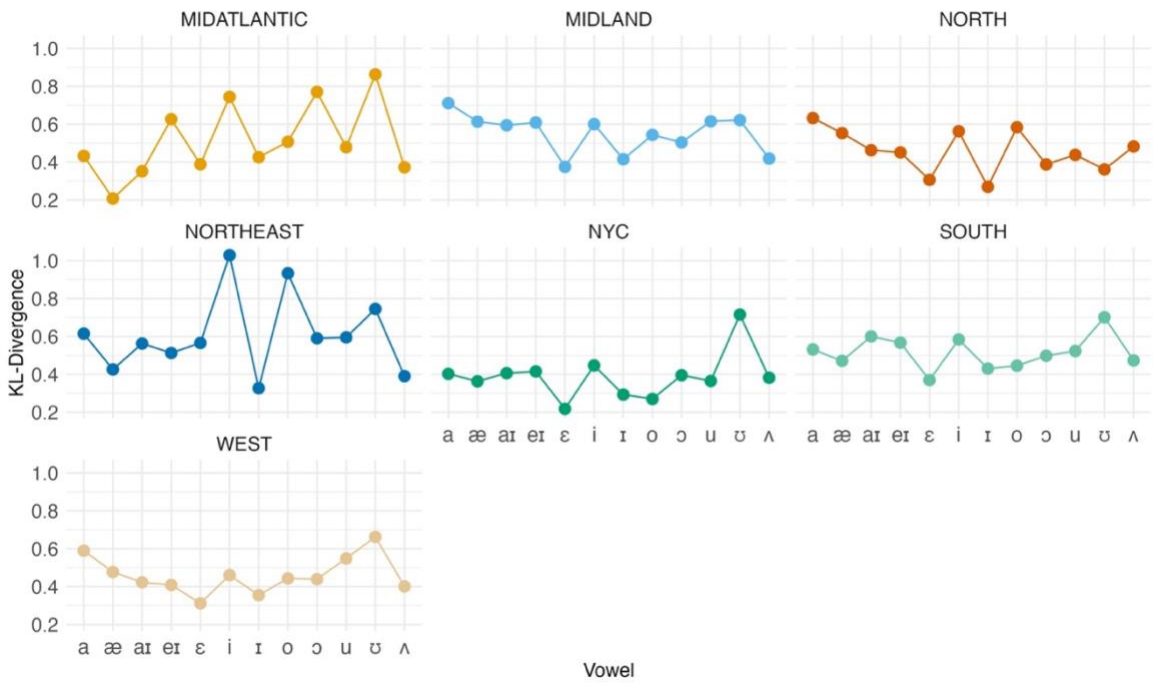

Figure 4.6 Mean KL divergence for the Talker factor from their dialect areas' distributions. Averaged over individual talkers faceted by vowel and regional group.

Table 4.7 Mean KL divergence for the Talker factor from their dialect areas' distributions. Averaged over individual talker levels faceted by vowel and regional dialect, rank ordered by vowel within dialect areas.

Northeast		**South**		**West**	
Vowel	**Mean KLD**	**Vowel**	**Mean KLD**	**Vowel**	**Mean KLD**
i	1.03	ʊ	0.70	ʊ	0.66
o	0.93	aɪ	0.60	a	0.59
ʊ	0.75	i	0.58	u	0.55
a	0.61	eɪ	0.57	æ	0.48
u	0.60	a	0.53	i	0.46
ɔ	0.59	u	0.52	o	0.44
ɛ	0.57	ɔ	0.50	ɔ	0.44
aɪ	0.56	æ	0.47	aɪ	0.42
eɪ	0.51	ʌ	0.47	eɪ	0.41
æ	0.43	o	0.45	ʌ	0.40
ʌ	0.39	ɪ	0.43	ɪ	0.36
ɪ	0.33	ɛ	0.37	ɛ	0.31
Mean	**0.61**	**Mean**	**0.52**	**Mean**	**0.46**

Table 4.7, Continued

Midland		Midatlantic		North	
Vowel	**Mean KLD**	**Vowel**	**Mean KLD**	**Vowel**	**Mean KLD**
a	0.71	ʊ	0.86	a	0.63
u	0.62	ɔ	0.77	o	0.58
ʊ	0.62	i	0.74	i	0.56
æ	0.61	eɪ	0.63	æ	0.55
eɪ	0.61	o	0.51	ʌ	0.48
i	0.6	u	0.48	aɪ	0.46
aɪ	0.59	a	0.43	eɪ	0.45
o	0.54	ɪ	0.43	u	0.44
ɔ	0.5	ɛ	0.39	ɔ	0.39
ʌ	0.42	ʌ	0.37	ʊ	0.36
ɪ	0.41	aɪ	0.35	ɛ	0.31
ɛ	0.38	æ	0.21	ɪ	0.27
Mean	**0.55**	**Mean**	**0.51**	**Mean**	**0.61**

NYC	
Vowel	**Mean KLD**
ʊ	0.72
i	0.45
eɪ	0.42
aɪ	0.41
ɔ	0.40
a	0.40
\	0.38
u	0.37
æ	0.36
ɪ	0.29
o	0.27
ɛ	0.22
Mean	**0.61**

4.1.7 Interim Summary

Overall, this analysis provides evidence of dialectal informativity of vowel categories both within and across regions. The results of this analysis are in line with the higher-order levels of informativity indicated in Kleinschmidt (2019) whereby more granular socio-indexical factors (e.g., Talker) provide more information than broader factors (e.g., Gender). Within a Dialect-

agnostic perspective (Section 4.1.3.), individual vowel categories demonstrating higher Dialect informativity across regions are /ɔ/, /aɪ/, and /eɪ/. For the factor of Talker, individual vowel categories demonstrate Talkers are most informative of the cue distributions for /ʊ/, /ʌ/, and /ɪ/. The asymmetry in this ranking suggests some vowel categories are broadly Dialect-informative while others are Talker-informative. Such a pattern is further supported by the fact that Talker informativity remains highest for the same categories when estimating talkers by their own regional distributions (i.e., nested structure), rather than the broader marginal distribution. Similarly, the Dialect-informative vowels are ranked lower in Talker informativity when using the nested structure.

The expectation that Dialect-agnostic patterns would align with categories implicated across regional vowel shifts was only weakly evident, as /aɪ/ and /eɪ/ were not generally expected to be conditioned strongly on Dialect. Consequently, a Dialect-specific perspective (Section 4.1.4) demonstrated that vocalic informativity within regions may be weakly linked to regional vocalic variability. When examining the patterns in terms of the more nested structure (Section 4.1.5) we see individual variability in vowel categories related to regional shifts. However, we still see the categories most highly ranked in terms of talker informativity are those associated with high talker informativity in general and are not strongly associated with regional patterns (e.g., /ʊ/).

4.2 Analysis 2: Shift Based Regions as Baseline

For this analysis, I go on to ask the same set of questions but with a smaller subset of the data presented in the previous section, focusing only on three dialect areas which are most representative of widescale vowel shifts (South, North, West) and drawing on only data from the Switchboard Corpus. This section aims to validate the observations made in the previous analysis on a more controlled subsample of the data. Given the observations in the previous section did not align with expectations around regional vowel shifts, examining a dataset focused on the more relevant regional areas provides an opportunity to examine regional dialects with fewer dimensions of variability to assess the validity of the results. While this method is less ecologically valid in terms of variability listeners are exposed to, it may have the benefit of aligning with the observations made in speaker perception whereby listeners predominately

categorize speakers into these three regional dialect areas (e.g., Clopper & Pisoni, 2007; Clopper et al., 2006).

4.2.1 Data & Method

To address this question, I analyze a subset of the data from a single dataset, the Switchboard corpus (see Chapter 3), and a subset of the data composed of three dialect areas: North, South, and West, associated with prominent regional vowel shifts NCS, SVS, and LBMS respectively. The data are thus comprised of the following factors: the marginal distribution $Q_S(\text{F1xF2} \mid \text{vowel, Switchboard[N,S,W]})$, Dialect factor with 3 levels, $D(Q_S||P_{di})$, gender with 2 levels, $D(Q_S||P_{gi})$, and Dialect+Gender with 6 levels (3 dialect levels x 2 gender levels) $D(Q_S||P_{dgi})$, and Talker with 146 levels (i.e., individuals), $D(Q_S||P_{Ti})$. Following the structure of the previous analysis in Section 4.1, the first subsections (Section 4.2.2-4.2.4) will evaluate KL divergence for each socio-indexical factor in relation to the 'baseline' marginal distribution. The marginal distribution in these subsections is made-up of all tokens from all talkers from the three dialect areas of the switchboard dataset (Q_S). I will then turn to look at a 'nested' version where talkers (P_{Ti}) are estimated in reference to their own dialect areas (Q_{Di}, Section 4.2.5), where the marginal distribution is the talkers own regional background limited to the Switchboard corpus. The number of unique speakers for each factor and their respective levels are provided in Table 4.8 and total token counts for the marginal distribution of each vowel category is presented in Table 4.9.

Table 4.8 Total talker counts by socio-indexical factor for the subset data: Switchboard data for North, South, West

Dialect	Men (N)	Women (N)	Dialect (N)
North	37	21	58
South	22	16	38
West	29	21	50
Total	**88**	**58**	**146**

Table 4.9 Total token counts for the marginal distribution by vowel categories for the subset data: Switchboard data for North, South, West.

Vowel	N
a	4043
æ	4172
ʌ	6501
ɔ	5084
aɪ	8148
ɛ	7764
eɪ	6565
i	6058
ɪ	4118
o	7934
u	3996
ʊ	1105
Total	**68216**

4.2.2 Higher-Order Factors

Figure 4.7 and Table 4.10 illustrate KL divergence across vowel categories and socio-indexical factors over the marginal distribution, as well as effects from a random assignment of individuals into groups, here with 2 groups (equal to Gender in the study) of randomly shuffled speakers ('Random-Gender') and 3 groups equal to the number of dialect groups ('Random-Dialect'). Again, following Kleinschmidt (2019) and the results in the section above, we observe that the more specific the grouping factor becomes (e.g., gender → talker) the more informative the factor is. Gender (mean KL = 0.02 bits) overall is the least informative factor across vowel categories, followed by Dialect (mean KL = 0.05 bits), and then by Talker (mean KL = 0.81 bits) as the most informative across vowel categories. Gender appears as informative as a random assignment of talkers into two groups across vowel categories, meaning there is not much conditioning on the normalized vowels by Gender alone. Dialect is slightly more informative than a random assignment of talkers into three groups, but across several categories remains on par with the Random-Gender factor. Talker is more informative than a random assignment of tokens within phonemes to talkers across vowel categories. Overall, this confirms the results

above that more specific groups provide more information over the broader grouping factors. Following the previous analysis, I will again refer to the respective rank ordering between Dialects and Talkers in the following sections.

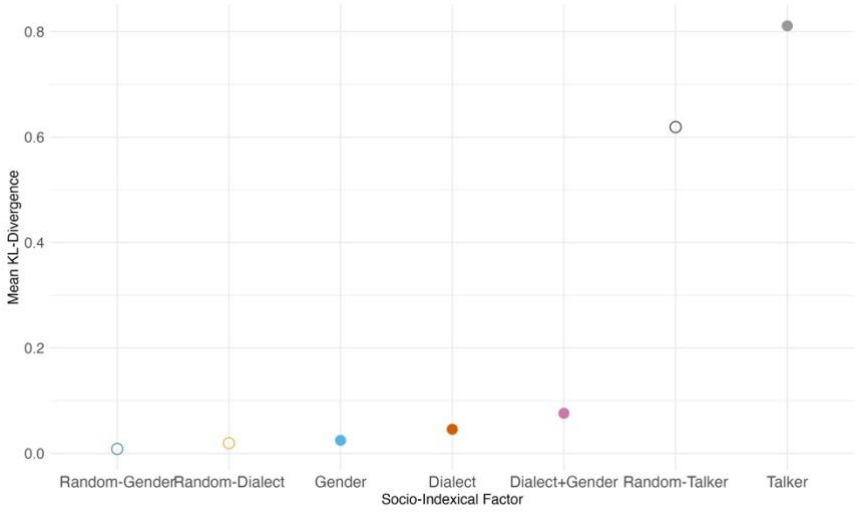

Figure 4.7 Mean KL divergence for each socio-indexical factor (filled circles), including randomly assigned talkers (& tokens) to comparable sized groupings (unfilled circles). Averaged over respective levels and vowel categories.

Table 4.10 Mean KL divergence for each socio-indexical factor and randomized groups.

Factor	Mean KL
Talker	0.81
Random-Talker	0.62
Dialect+Gender	0.08
Dialect	0.05
Random-Dialect+Gender	0.04
Gender	0.02
Random-Dialect	0.02
Random-Gender	0.01

4.2.3 Vowel-Specific: Dialect-Agnostic & Talkers

Figure 4.8 and Table 4.11 provides mean KL divergence values across dialect areas and talkers, respectively, rank ordered. In line with Section 4.1, the rank orderings for Dialect and Talker are inversely related, confirming the asymmetry above that Dialect is informative of some vowel categories' cue distributions and Talker is highly informative of other vowel categories. A familiar pattern emerges where Talker is highly informative of /ʊ/ cue distributions (1.66 bits), in addition to /u/ (1.16 bits) and /æ/ (1.12 bits) cue distributions. A random assignment of /ʊ/ tokens to talkers (Random-Talker) are higher in informativity than the true Talker factor, which suggests it may not be a stable effect in this dataset and may be linked to smaller variance distributions. Dialect demonstrates higher informativity of /aɪ/ (0.09 bits), and /eɪ/ (0.09 bits) distributions which is in-line with the observations from Section 4.1 but once again are not reflective of the categories we would predict based on regional shifts. Additionally, it should be noted that the KL divergence values are generally lower for the Dialect factor across vowel categories in this subset of data compared to Section 4.1 and does not differ from a random grouping across most vowel categories, except for /aɪ/ and /eɪ/. The lower informativity by Dialect could be caused by several factors. The first is simply the result of a mathematical leveling, whereby each group represents approximately one-third of the marginal distribution and is thus well estimated by the marginal and does not comprise many low probability examples. This is potentially likely given the random assignment groups, at least in the aggregate. Another potential explanation, which I explore next, is that higher informativity can be found in dialect-specific and vowel-specific combinations, which is not reflected in the aggregate patterns here.

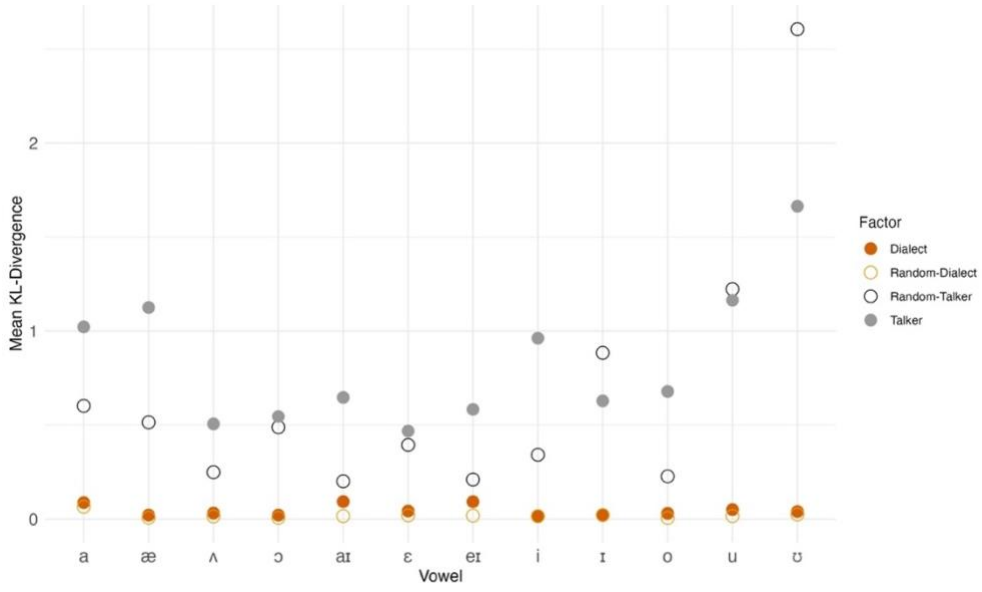

Figure 4.8 Mean KL divergence for the socio-indexical factors of Talker and Dialect (filled circles), including randomly assigned talkers (& tokens to talkers) to comparable sized groupings (unfilled circles), over marginal distribution (shifted regions). Averaged over respective levels and separated by vowel category.

Table 4.11: Mean KL divergence for the socio-indexical factors of Talker and Dialect over marginal (shifted regions) distributions, rank ordered respectively.

Dialect		**Talker**	
Vowel	**Mean KLD**	**Vowel**	**Mean KLD**
a	0.09	ʊ	1.66
aɪ	0.09	u	1.16
eɪ	0.09	æ	1.12
u	0.05	a	1.02
ɛ	0.04	i	0.96
ʊ	0.04	o	0.68
ʌ	0.03	aɪ	0.65
o	0.03	ɪ	0.63
æ	0.02	eɪ	0.58
ɔ	0.02	ɔ	0.55
i	0.02	ʌ	0.51
ɪ	0.02	ɛ	0.47
Mean	**0.05**	**Mean**	**0.83**

4.2.4 Vowel-Specific: Dialect-Specific

In this section, I turn to examine how structure emerges in dialect-specific and vowel-specific ways. Figure 4.9 illustrates KL divergence for each vowel category by each dialect area. First, there is a general observation that the North and South are more informative of vowel distributions on average than the West, as was indicated in Analysis 1 as well. Specific dialect levels vary in how informative they are about individual vowel categories. Such patterns further align with some expectations of regional vocalic shifts. In particular, we see the South is informative of cue distributions for /aɪ/ (KL = 0.17 bits), /eɪ/ (KL = 0.20 bits), and /a/ (KL = 0.12 bits), aligning with Analysis 1 and expectations for the SVS more broadly. The North demonstrates higher informativity of /a/ (KL = 0.10 bits) and /aɪ/ (KL = 0.08 bits), again aligning with the observations in Analysis 1, albeit with generally lower values. Finally, the West generally has the lowest information gained from the marginal distribution, but the ranking of vowel categories aligns with the patterns above (Section 4.1), with the highest category as /eɪ/ (KL = 0.05 bits), followed by /u/ (KL = 0.05 bits). Overall, the dialect-specific and vowel-specific patterns in this section generally align with some expectations of regional vowel shifts, though not overwhelmingly for the North or West, and aligning with the results of Analysis 1. Further, the recurrence of the ranking of the West as lower in informativity to cue distributions in Analysis 1 and Analysis 2 aligns with listeners' subjective evaluations of more unmarked speech patterns. The average informativity of the North and South aligns with listener evaluations of marked speech patterns, despite the fact that expected vowel categories do not necessarily emerge as the most informative. Of course, once again, the reason certain vowel categories do not emerge in the North and South may be driven by talker variability within the region for particular categories, a point I will turn to in the next section.

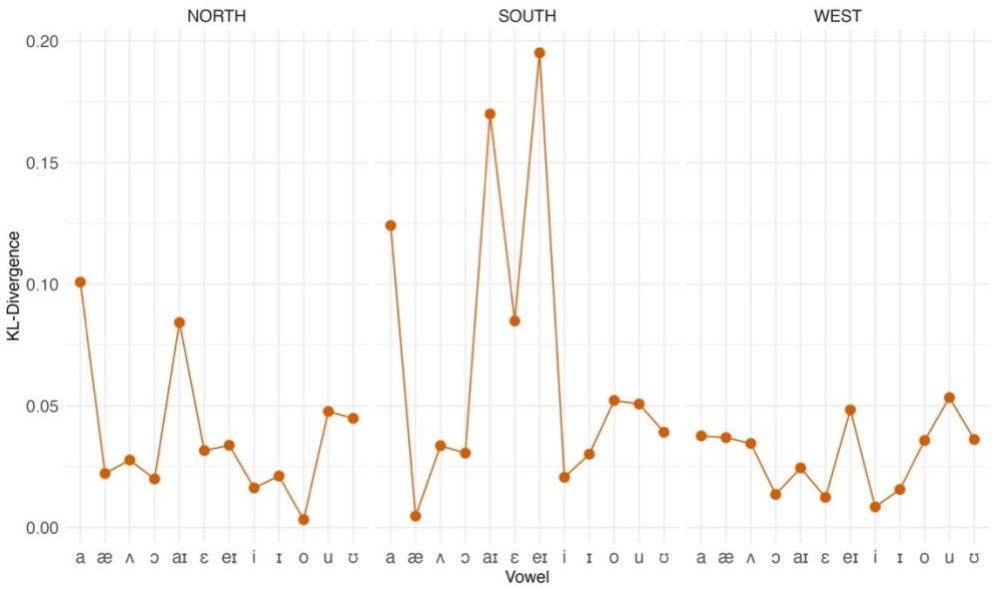

Figure 4.9 KL divergence values each level of the Dialect factor (i.e., individual dialect areas) by vowel category.

Table 4.12 KL divergence values each level of the Dialect factor (i.e., individual dialect areas) by vowel category, rank ordered within regions.

North		South		West	
Vowel	**Mean KLD**	**Vowel**	**Mean KLD**	**Vowel**	**Mean KLD**
a	0.10	eɪ	0.20	eɪ	0.05
aɪ	0.08	aɪ	0.17	u	0.05
u	0.05	a	0.12	a	0.04
ʊ	0.04	ɛ	0.08	æ	0.04
ʌ	0.03	o	0.05	o	0.04
ɛ	0.03	u	0.05	ʊ	0.04
eɪ	0.03	ʊ	0.04	ʌ	0.03
æ	0.02	ʌ	0.03	aɪ	0.02
ɔ	0.02	ɔ	0.03	ɪ	0.02
i	0.02	ɪ	0.03	ɔ	0.01
ɪ	0.02	i	0.02	ɛ	0.01
o	0.00	æ	0.00	i	0.01
Mean	**0.04**	**Mean**	**0.07**	**Mean**	**0.03**

4.2.5 Analysis 2b: Talkers Within Dialects (Nested)

To proceed, I again turn to the question of how well individual distributions align with their dialect areas by examining a more nested structure to evaluate talker divergence from their dialect areas. In this section, the marginal distribution is now individual talkers' own dialect regions, $Q_{MDi}(_{F1xF2 \mid vowel,di})$, where KL divergence is measured by evaluating how much information is to be gained from talker-specific distributions (P_{Ti}) over and above their dialect area estimates. As such, the data in this section report new values from those of Sections 4.2.2-4.2.4 and shed light on the extent to which talkers within dialect areas adhere to their group patterns. First, we see that on average Talker informativity is lower (mean = 0.48 bits) when using their own dialect areas to estimate their cue distributions compared to the broader marginal in Section 4.2.2-4.2.4. This is in-line with Analysis 1 and validates that estimating talkers by their regional dialects results in less information loss providing support for the hypothesis that talkers tend to align with their group's distributional properties. Turning to vowel-specific patterns in Table 4.13 we see confirmation of the patterns observed in Section 4.2.4, where Talker is most informative of the vowel categories that ranked higher in Talker informativity using the broad marginal distribution. Talker is more informative of categories like /ʊ/ and /a/ even when talkers' dialect areas are used to estimate individual talker distributions. This finding is further confirmation of an asymmetry, at least for the dialect-agnostic perspective that some categories are Dialect-informative while others are Talker-informative, showing limited dialect conditioning. Further, the Dialect-informative categories generally show less information loss when estimating individual talkers' categories with their dialect area distributions.

Table 4.13 Mean KL divergence for the Talker factor from their dialect areas' distributions (subset group), rank ordered. Averaged over individual talkers across regional backgrounds.

Vowel	Mean KL
a	0.61
aɪ	0.50
eɪ	0.47
i	0.60
o	0.52
u	0.48

Table 4.13 Continued

æ	0.55
ɔ	0.38
ɛ	0.32
ɪ	0.34
ʊ	0.54
ʌ	0.43
Mean	**0.48**

4.2.6 Talker-Specific Patterns by Dialect

Before turning to vowel specific patterns, we can see that across vowel categories in Table 4.14, Talker information is highest for the South (mean KL = 0.54 bits), followed by the North (mean KL = 0.47 bits) and the West (mean KL = 0.44 bits) which aligns with Analysis 1 for the same regions. Looking at dialect-specific patterns, Table 4.14 and Figure 4.10 shows the average KL Divergence across talkers for each vowel category by dialect area. Overall, there is weak alignment with the hypothesis that categories more canonically associated with the region may demonstrate greater Talker informativity (i.e., higher KL divergence). In particular, for the South, Talker is informative of /i/ (mean KL = 0.83 bits), and to some extent /aɪ/ (mean KL = 0.64 bits) and /eɪ/ (mean KL = 0.61 bits), in line with Analysis 1 and more general expectations of reversal of the SVS and more geographic restriction of shifted /i/ across the region. However, this departs from the previous analysis in that the South was also high in informativity for these categories but is simultaneously showing greater within-region between-talker variability. This may illustrate that the South on average shows greater divergence from the marginal for these categories and a wide range of inter-talker variability. Such a pattern challenges the perspective outlined in Kleinschmidt (2019) where talkers generally mirror their dialect areas. In the North we see a parallel pattern where Talker is more informative of /a/ (mean KL = 0.60 bits) distributions, showing greater talker divergence on average despite greater dialect-specific informativity for /a/ in Section 4.2.4. Finally, the West demonstrates Talker information is highest for /a/ (mean KL = 0.60 bits) and /æ/ (mean KL = 0.53 bits), which aligns generally with both categories being integral to the LBMS, with /æ/ showing retraction and /a/ being involved in the low-back vowel merger.

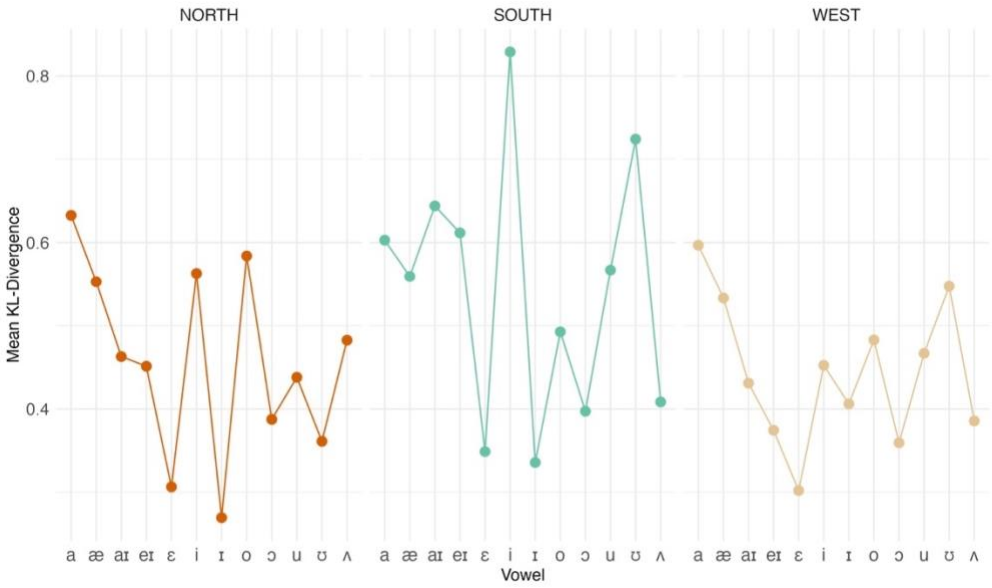

Figure 4.10 Mean KL divergence for the Talker factor from their dialect areas' distributions. Averaged over individual talker levels faceted by vowel and regional group.

Table 4.14 Mean KL divergence for the Talker factor from their dialect areas' distributions. Averaged over individual talker levels faceted by vowel and regional dialect, rank ordered by vowel within dialect areas.

South		West		North	
Vowel	**Mean KLD**	**Vowel**	**Mean KLD**	**Vowel**	**Mean KLD**
i	0.83	a	0.60	a	0.63
ʊ	0.72	ʊ	0.55	o	0.58
aɪ	0.64	æ	0.53	i	0.56
eɪ	0.61	o	0.48	æ	0.55
a	0.60	u	0.47	ʌ	0.48
u	0.57	i	0.45	aɪ	0.46
æ	0.56	aɪ	0.43	eɪ	0.45
o	0.49	ɪ	0.41	u	0.44
ʌ	0.41	ʌ	0.39	ɔ	0.39
ɔ	0.40	eɪ	0.37	ʊ	0.36
ɛ	0.35	ɔ	0.36	ɛ	0.31
ɪ	0.34	ɛ	0.30	ɪ	0.27
Mean	**0.54**	**Mean**	**0.45**	**Mean**	**0.46**

4.2.7 Interim Summary

Analysis 2 largely confirms the results in Analysis 1, albeit with lower overall values for Dialect informativity. In particular, Analysis 2 demonstrated that from a dialect-agnostic perspective, Dialects on average are most informative of /a/, /aɪ/ and /eɪ/ cue distributions and Talker is most informative of /ʊ/ and /u/ cue distributions. Talkers diverge least from their dialect areas for categories that emerge as, broadly, dialect-informative in this dataset (e.g., /eɪ/). The asymmetry between the Dialect ranking and Talker ranking of categories again supports the fact that some vowel categories may be broadly Dialect-informative while others are Talker-informative. When we examine dialect-specific patterns, we see that individual dialect areas vary in terms of how informative they are of the vowel space on average (i.e., across vowel categories) and of specific vowel categories. In particular, in Analysis 2 we see that the North and South demonstrate higher informativity across vowel categories compared to the West, which shows lower values of informativity in relation to the marginal. Overall, however, we generally see that the North and South do not emerge as particularly informative of cue distributions for categories most robustly associated with vowel shifts associated with the respective regions. These results may be explained by the fact that Talkers within the region are more variable across vowel categories highly salient within the region, though this is only weakly supported by the data (Section 4.2.5).

An alternative explanation is that categories that are more likely to show low-level phonetic uniformity across talkers within regions may not necessarily be those most canonically associated with regional varieties. Given the results in these two analyses, we do see dialect-informative categories emerge from the dialect-agnostic perspective that suggests talkers within regions may be more uniform in their cue distributions, including /eɪ/ and /a/. Such low-level variation may not be hypothesized to be socially salient but may prove informative to listeners for both social and linguistic perception. Some evidence for this perspective can be drawn from Gunter et al. (2020) where the vowel categories demonstrating the most regularity across talkers of Southern identity were evaluated as most accented, despite the fact that the categories were not hypothesized to be saliently associated with the South. From the data here, we might predict that dialect-informativity for specific vowel categories across talkers may be worth for adaptation. However, it remains unclear whether such tracking occurs in a dialect-agnostic way,

aggregating over experiences with dialects, dialect-specific ways, or some combination of the two. Furthermore, given the high degree of talker-specific information of vowels (and the magnitude more information gained by Talker), whether such adaptation is sensitive to group-specific patterns at all remains to be seen. I will revisit some of these questions in more depth in the Discussion (Section 5) and outline more specific hypotheses.

Table 4.15: Summary of top ranked informativity of socio-indexical component by vowel categories across Analysis 1 and 2. Grey check marks indicate highly informative of socio-indexical component in only one analysis, black indicates both analyses.

Vowel	Dialect-agnostic	Talker	Talker (Nested)
a	✓		
ʊ		✓	✓
æ			
o			
u		✓	
i			✓
aɪ	✓		
ɪ		✓	
ʌ		✓	
eɪ	✓		
ɔ	✓		
ɛ			

4.3 Analysis 3: Single Region Baseline

In this section, I move on to consider a different type of 'baseline' experience than the previous sections have employed. In Sections 4.1 and 4.2, the marginal distribution was an overall depiction of American English, acting as a simulation for which we consider listeners' priors. However, such a sample assumes that the prior from which listeners infer socio-indexical structure is a broad and diverse sample of American English. Yet, it's reasonable to assume that listeners may be gleaning information relative to their own dialect areas rather than the entirety of American English. Listeners may be likely to approach speech processing tasks with representative speech from their community and the information gained by knowing dialect or

talker information may be a function of how dissimilar the speech is from their own regional background. Additionally, it's reasonable to expect that individual vowel categories more likely to be described by differences between dialect areas or an idealized average (i.e., benchmarking as in Labov et al., 2006) rather than differences from an aggregate of American English more broadly. To simulate this perspective, in this section I use only one region, the West, as the 'baseline' experience and evaluate KL divergence of Talkers and Dialects from the West. The West was chosen because it comprises a statistically sound amount of data in the datasets and is a convenient starting point for comparisons.

4.3.1 Data & Method

The data from this section include the entire dataset across all regions and groups presented in Chapter 3. The marginal distribution in this section assumes all of the data comprising the 'West' to represent a single community baseline experience. Following the methods in Sections 4.1- 4.2, the Talker factor represents an average of all individual talkers, and the Dialect factor represents an average of all dialect levels. When relevant, individual levels of the factors will be explicitly indicated in the figures and prose. The data are thus comprised of the following factors: the reference distribution $Q_W(\text{F1xF2} \mid \text{vowel, West})$, Dialect factor with 3 levels, $D(Q_W||P_{di})$, gender with 2 levels, $D(Q_W||P_{gi})$, and Dialect+Gender with 6 levels (3 dialect levels x 2 gender levels) $D(Q_W||P_{dgi})$, and Talker with 146 levels (i.e., individuals), $D(Q_W||P_{Ti})$. The counts for the reference distribution for each vowel category are provided in Table 4.16. The respective breakdown of unique talkers in each dialect area is the same as Analysis 1, but for convenience has been replicated in Table 4.17 below. Diverging from the previous analyses, I am not representing a random assignment of talkers to dialect areas, as we would largely expect that any arrangement of talkers from variable different backgrounds is still likely to diverge from a single dialect area and is, therefore, less useful in capturing a null result. Additionally, for this analysis, it is worth noting that for the Dialect factor analysis, the reference distribution (Q_W) is comprised of a smaller sample than some of the other dialect areas (i.e., P_{Di}*:* Northeast, Midland, and South). Given KL divergence is asymmetrical, we should expect that a smaller distribution will be less optimal for encoding distributions that may be more variable simply as a function of the larger number of unique speakers in the sampled *true* distribution (i.e., Dialect areas).

However, as the results will show, the effects remain similar at the Talker level where such an asymmetry is not apparent, suggesting the dialect pattern may be representative of overall differences between talkers across regional dialects rather than the asymmetrical sample sizes alone.

Table 4.16 Total token counts for the reference distribution by vowel category for a single dialect region, the West.

Vowel	N
a	5477
æ	6443
ʌ	7083
ɔ	6134
aɪ	12965
ɛ	10376
eɪ	9097
i	9263
ɪ	6194
o	11231
u	4908
ʊ	855
Total N	**90428**

Table 4.17 Total talker counts for each socio-indexical factor, replicated from Analysis 1. Note, the total talkers in the reference distribution equals 120, which is lower than several of the compared distributions (e.g., South, Midland, etc.)

Dialect Area	Men (N)	Women (N)	Dialect (N)
Midatlantic	5	9	14
Midland	153	151	304
North	37	21	58
Northeast	108	108	216
NYC	11	8	19
South	109	103	212
West	58	59	117
Total N	**481**	**459**	**940**

4.3.2 Higher-Order Factors

Figure 4.11 illustrates the higher-order factor averages which have been averaged again across vowel categories. Once again, the figure confirms that socio-indexical factors provide information about cue distributions but differ in magnitude, and more specific groupings are more informative of cue distributions. In line with previous analyses and Kleinschmidt (2019), Gender provides little information gained for the normalized distributions on average (0.06 bits), followed by an increase in information gained by Dialect (0.14 bits), and finally Talker (1.22 bits). The magnitude and rank ordering of each socio-indexical factor generally aligns with Analysis 1 despite rather distinct reference distributions, where Analysis 1 contained all data, but only a single dialect area is used here. Together, the findings of Analysis 1-3 establish stable information gained across different socio-indexical factors.

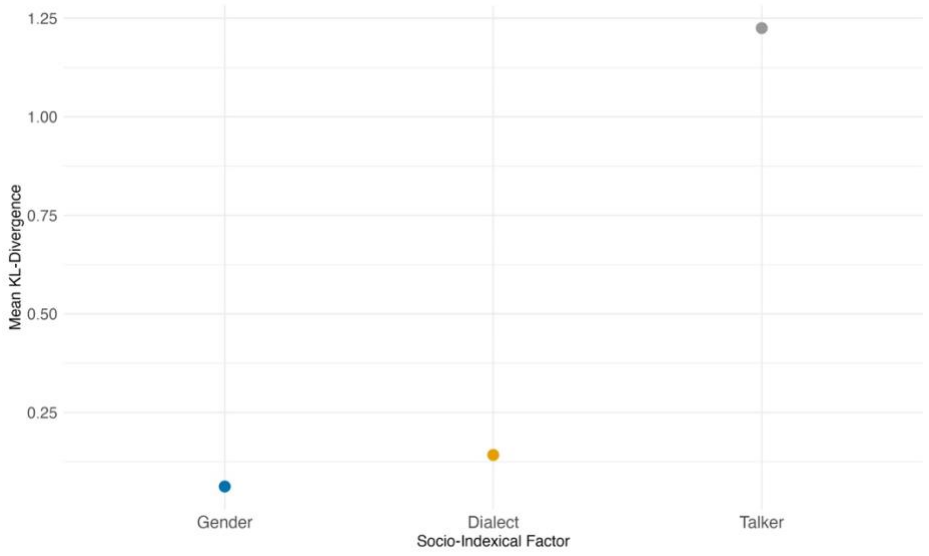

Figure 4.11 Mean KL divergence for each socio-indexical factor, averaged over respective levels and vowel categories.

Table 4.18 Higher order factors KL Divergence

Factor	KL
Talker	1.22
Random-Talker	0.71
Dialect+Gender	0.17
Dialect	0.14
Random-Dialect+Gender	0.10
Random-Dialect	0.08
Gender	0.06
Random-Gender	0.05

4.3.3 Vowel-Specific: Dialect-Agnostic and Talkers

Figure 4.12 shows both the Dialect and Talker factors for each vowel category, with Dialect and Talker factors represented as averages over the individual levels conditioned on each vowel category. Figure 4.12 illustrates vowel-specific trends of each factor that align, to some extent, with the previous analyses. Again, we see Dialect emerges as most informative of /a/ (0.24 bits) distributions followed by /aɪ/ (0.21 bits). Dialect emerges as nearly equal in informativity for /æ/ (0.21 bits), /ɔ/ (0.20 bits), and /eɪ/ (0.19 bits). Unsurprisingly, the values for Dialect are in general higher across these categories than the previous analyses, which is to be expected given the smaller and less representative sample acting as the reference group. Overall, some categories appear to be reliably informative in dialect-agnostic ways, as evidenced by the stability of several of these categories emerging as dialect-informative across analyses. This corroborates that dialects provide informativity to vowel categories that are not necessarily associated with more salient patterns among dialect areas, like /a/ and /ɔ/. Departing from previous analyses, however, we see that /æ/ seems to rank higher in informativity, aligning with expectations based on previous work across vowel shifts. Though, as in the previous analyses here, categories like /ɪ/ and /ɛ/ generally do not emerge as Dialectally informative.

Turning our attention to the Talker factor, Figure 4.12 and 4.18 alignment with results from Analysis 1 and 2. For example, Talkers emerge as most informative of /ʊ/ (and Dialect is least informative of /ʊ/). Otherwise, there is some deviation from previous analyses, as we see /ʌ/ and /ɪ/ are Talker informative here. We also see that /eɪ/ and /a/ generally emerge as somewhat highly ranked for Talker information, despite the fact that they emerge as more highly ranked by Dialect. Of course, this not surprising given the fact that all talkers in this analysis are from other regions outside of the West. Thus, it makes sense that Talker emerges as highly informative of the same vowel categories as dialect areas since all talkers are from different dialect areas from the reference distribution.

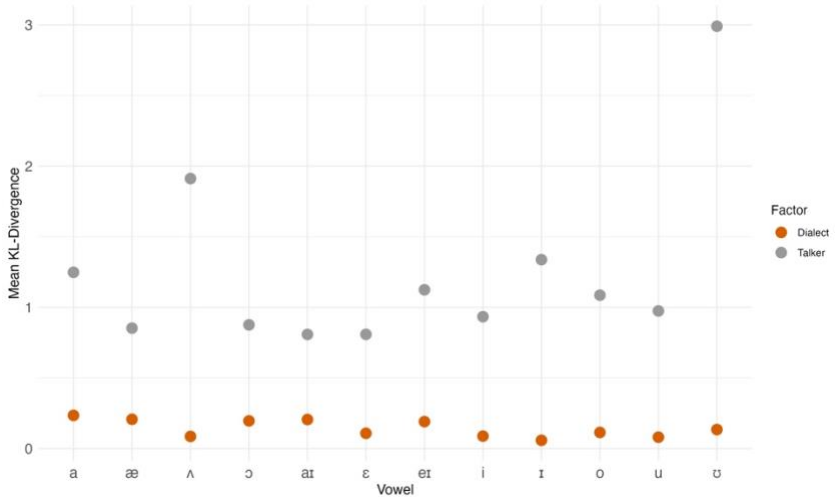

Figure 4.12 Mean KL divergence for the socio-indexical factors of Talker and Dialect (filled circles). Averaged over respective levels and separated by vowel category.

Table 4.19 Mean KL divergence for the socio-indexical factors of Talker and Dialect over single region baseline distributions, rank ordered respectively.

Dialect		**Talker**	
Vowel	**KL**	**Vowel**	**KL**
a	0.24	ʊ	2.99
æ	0.21	ʌ	1.91
aɪ	0.21	ɪ	1.34
ɔ	0.20	a	1.25
eɪ	0.19	eɪ	1.12
ʊ	0.13	o	1.09
ɛ	0.11	u	0.97
o	0.11	i	0.93
ʌ	0.09	ɔ	0.88
i	0.09	æ	0.85
u	0.08	aɪ	0.81
ɪ	0.06	ɛ	0.81
Mean	**0.14**	**Mean**	**1.22**

4.3.4 Vowel-Specific: Dialect-Specific

Figure 4.13 and Table 4.20 present the KL divergence values for individual dialect levels and vowel categories, rank ordered within regions. First, generally speaking, each dialect area on average diverges from the West, with the Northeast (mean = 0.24 bits) and South (mean = 0.19 bits) demonstrating the highest divergence. On the other hand, the Midland area demonstrates very little divergence from the West (mean = 0.07 bits). This result reaffirms the previous analyses and aligns with listeners' perception of speech in the West and Midland area as predominately unmarked (Preston, 2011), and the speech of the South and Northern areas as marked (Preston, 1996).

Looking more granularly, the results in this section depict patterns aligning more closely with expectations from sociophonetic literature than seen in the previous two analyses (Section

4.1 – 4.2). In particular, we see that the North and Northeast are most informative of the cue distributions for /a/ (0.32 and 0.74 bits respectively) and /æ/ (0.37 and 0.46 bits) and with the Northeast showing clear distinction in /ɔ/ (0.52 bits). This is unsurprising given the preservation of the low-back vowels in both areas compared to the West, and that /a/ and /æ/ occupy different positions than the West. The South provides information about /ɔ/ distributions (0.23 bits), though does not provide the same degree of detail about /a/ (0.03 bits) as the North and Northeast. The South, however, does appear to be the most informative of categories canonically associated with the SVS, including /eɪ/ (0.48 bits), /aɪ/ (0.60 bits), and /ɛ/ (0.23 bits). This pattern aligns rather well with listeners' social perception of Southern speech (Albritten, 2011; Plichta, & Preston 2005) and descriptions of prominent vowel categories by researchers (e.g., Fridland, 2000 Thomas, 2001). Overall, the patterns in this section confirm expectations of informativity of individual dialect areas in comparison to the West as a baseline and generally align with existing descriptions of listener-oriented behavior. Whether the same categories or granularity operate for listeners' inferences and linguistic categorization behavior is less clear, as I will discuss in more detail in Section 5.

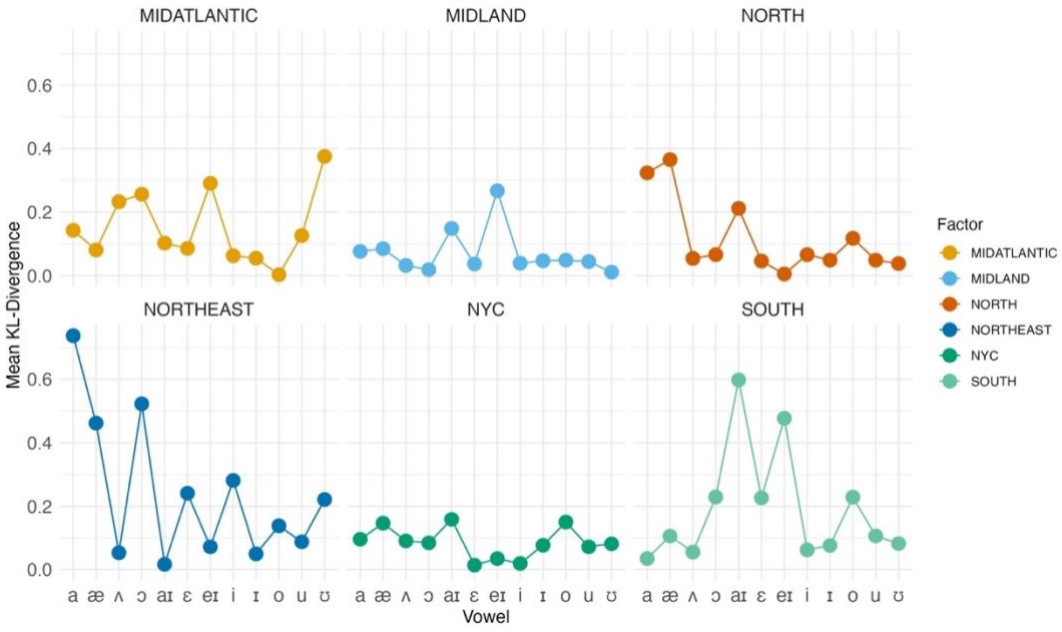

Figure 4.13 KL divergence values each level of the Dialect factor (i.e., individual dialect areas) by vowel category.

Table 4.20 KL divergence values each level of the Dialect factor (i.e., individual dialect areas) by vowel category, rank ordered within regions.

Midatlantic		Midland		North		Northeast	
Vowel	**KL**	**Vowel**	**KL**	**Vowel**	**KL**	**Vowel**	**KL**
ʊ	0.38	eɪ	0.27	æ	0.37	a	0.74
eɪ	0.29	aɪ	0.15	a	0.32	ɔ	0.52
ɔ	0.26	a	0.08	aɪ	0.21	æ	0.46
ʌ	0.23	æ	0.08	o	0.12	i	0.28
a	0.14	ɪ	0.05	ɔ	0.07	ɛ	0.24
u	0.13	o	0.05	i	0.07	ʊ	0.22
aɪ	0.10	ɛ	0.04	ʌ	0.05	o	0.14
ɛ	0.09	i	0.04	ɛ	0.05	u	0.09
æ	0.08	u	0.04	ɪ	0.05	eɪ	0.07
i	0.06	ʌ	0.03	u	0.05	ʌ	0.05
ɪ	0.05	ɔ	0.02	ʊ	0.04	ɪ	0.05
o	0.00	ʊ	0.01	eɪ	0.01	aɪ	0.02
Mean	**0.15**	**Mean**	**0.07**	**Mean**	**0.12**	**Mean**	**0.24**

NYC		South	
Vowel	**KL**	**Vowel**	**KL**
aɪ	0.16	aɪ	0.60
æ	0.15	eɪ	0.48
o	0.15	ɔ	0.23
a	0.1	ɛ	0.23
ʌ	0.09	o	0.23
ɔ	0.08	æ	0.11
ɪ	0.08	u	0.11
ʊ	0.08	ɪ	0.08
u	0.07	ʊ	0.08
eɪ	0.03	i	0.06
i	0.02	ʌ	0.05
ɛ	0.01	a	0.03
Mean	**0.09**	**Mean**	**0.19**

4.3.5 Interim Summary

The results of this section provide further confirmation of the results from Analysis 1 and 2, particularly with regard to the dialect-agnostic and talker-specific patterns therein. However, when using a single dialect area as a baseline, dialect-specific patterns align more robustly with expectations from previous work in sociophonetics. This fact suggests that evaluating structure in production may be most representative when examining differences between regions or a single baseline as in the case of benchmarking. That is, unsurprisingly, the large aggregate of American English as a reference may mask dialect-specific tendencies.

Table 4.20 summarizes the results from across the three analyses for the dialect-agnostic, talker, and talker (nested) socio-indexical levels. Across these three analyses, two categories, /a/ and /aɪ/, have consistently emerged as the most informative at the dialect-agnostic level. Similarly, /eɪ/ and /ɔ/ have emerged as the most informative across two analyses and ranked only slightly lower in the third analysis. Overall, this demonstrates a high degree of talker regularity for /a/, /aɪ/, /eɪ/, and /ɔ/ within dialect areas. Across analyses, talker identity emerges as informative consistently for several vowel categories which are complemented by low-ranked dialect informativity for the same categories. Additionally, for one of these categories, /ʊ/, individual talkers are most divergent from their dialect area distributions. This demonstrates that /ʊ/ distributions are largely talker specific and not reliably conditioned by dialect areas. Overall, these results challenge the assumption that all vowels may provide equal degrees of socio-indexical structure and rather different vowel categories demonstrate different behaviors across socio-indexical factors. In the following section, I will discuss these results and their implications in greater detail.

Table 4.21: Summary of top-ranked informativity of socio-indexical component by vowel categories across all analyses. Grey check marks indicate highly informative of socio-indexical components in only one or two analyses, black indicates all three analyses.

Vowel	Dialect-agnostic	Talker	Talker (Nested)
a	✓		
ʊ		✓	✓
æ			

Table 4.21, Continued

o			
u		✓	
i			✓
aɪ	✓		
ɪ		✓	
ʌ		✓	
eɪ	✓		
ɔ	✓		
ɛ			

5 Discussion

I will now turn to a broad discussion of the results from the three analyses combined. I will first provide an overall summary of the patterns that emerged across the analyses and how they compare to the findings of Kleinschmidt (2019) and listeners' social perceptions of variation. Following this summary, I will turn to consider how the results align with expectations of production, and what these results tell us about socio-indexical structure and regional variation in production. Then, I will turn to consider how the different perspectives outlined in this chapter provide implications for listeners' perceptual learning behavior across vowel categories. I will organize these discussions around the posited socio-indexical levels described throughout this chapter. The first level distinguishes vowel categories by the asymmetry of Dialect-informative and Talker-informative, focusing specifically on the dialect-agnostic perspective. The second level differentiates dialect-specific patterns from the dialect-agnostic perspective, detailing when and how they may contrast.

The results from Analysis 1-3 validate Kleinschmidt (2019) in terms of the overall ranking of socio-indexical factors in informativity. Overall, we see the same rank ordering from broader to more specific socio-indexical groups demonstrating lower to higher levels of informativity (gender < dialect < dialect + gender < talker). However, the vowel-specific patterns of the Talker factor and Dialect factor were not an exact replication of the results presented by Kleinschmidt (2019). Yet, the dialect-informativity and talker-informativity of vowel categories are largely replicated across the analyses of this Chapter, despite not replicating the exact patterns in Kleinschmidt (2019). Specifically, the Dialect factor shows greater informativity

about /eɪ/ and /a/ distributions and the Talker factor greater informativity about /ʊ/ distributions. Information gained for Talker identity is also highest for /ʊ/ when the marginal distribution is the talkers' own dialect regions. Demonstrating that talkers are most divergent from their own dialects for /ʊ/ cue distributions. On the other hand, information gained by Talker identity is ranked lower for /eɪ/, suggesting that Talkers are on average similar to their dialect areas' distributions. These findings together provide some validation of the uniformity of talkers within social groups suggested by Kleinschmidt (2019), despite the fact that the vowel-specific trends do not necessarily align.

Examining several simulations of prior experience to evaluate informativity has yielded interesting insights, including that some vowels continually emerge as more strongly conditioned on dialect. This suggests that the model provides some relatively stable predictions from which we can test listener behaviors (as I'll discuss in detail below). Further, it highlighted some interesting patterns that supports more qualitative descriptions of listeners' evaluative perceptions of American English. In particular, when using the West as a 'baseline' region, the more salient categories emerge across regional dialects more robustly compared to using a broader conceptualization of American English. In addition, regions that were seen as more divergent from the broader 'American English' baseline also align with listener perceptions of salience. These findings overall provide some acoustic evidence of what listeners evaluate as relatively unmarked, or "general American" as largely aligning with the speech of the West and Midwestern region. Similarly, the greater divergence of specific regions is supported by listeners categorizing talkers into similar groups during regional categorization tasks (Clopper et al., 2006) and an overall shared understanding of how speech varies despite varied experiences with regional variation.

5.1 (Re)contextualizing Socio-Indexical Structure: Production

The results of this chapter provide initial insight into dialectal conditioning of distributions and how individuals' distributions align with their dialect areas and provide critical insight into structured variation. Over the course of the analyses above, the results highlight specific categories that may demonstrate the greatest regularity across talkers in a region, as demonstrated by lower divergence of talkers from their groups, and higher informativity of

Dialect as a factor. In previous work on vowel shifts we would expect categories that like /ɛ/ and /æ/ to demonstrate greater dialect-informativity due to their involvement across regional shifts (Labov et al., 2006, see Section 2.2 above). However, we don't see such a pattern across the aggregate factor of Dialect (i.e., dialect-agnostic) or robustly at the individual dialect levels (i.e., dialect-specific) consistently across the analyses. Rather, the categories most likely to be conditioned by dialects are /eɪ/ and the low back vowels /a/ and /ɔ/. Such findings provide insight into structured variation demonstrating that some categories may generally show an increased regularity among talkers, and they may not necessarily align with the most salient categories defining the dialect.

First, it is worth noting that some of the expected patterns from vowel shifts may not have emerged as dialect-informative because the formalization of socio-indexical structure in this chapter assumes a flat structure across a raw frequency distribution and does not consider more fine-grained linguistic conditioning. While this choice is motivated by previous work in ideal-adapter models and perceptual learning, it may have ramifications for the compatibility of sociophonetic work on vowel shifts. The internal conditioning of variation is at the core of much sociophonetic work, as variation is conditioned by a combination of socio-indexical factors and internal linguistic factors. As such, we might expect that taking internal linguistic factors into account may yield more alignment with expectations of regional patterns. This is one potential pitfall of the computational model as it is currently defined and warrants additional interrogation to fully integrate sociophonetic insights. While I will not address the full scope of internal linguistic conditioning, in Chapter 5 I will return to this point and examine relationships among vowels as one aspect of the internal constraints of variability. Nonetheless, the results provide additional insight into structured variation that are valuable to consider.

The information gained about cue distributions for the low-back vowels by Dialect is perhaps not surprising given that regions generally differ in terms of whether the two vowels are merged or not. Consequently, the distributional properties of the low back vowels are potentially greater than other categories because of apparent bimodality across American English. As such, we might expect that any one region will more likely have a reduced variance for these categories when compared to American English more broadly, where we would potentially see a unimodal distribution at the level of a dialect or individual talker. Due to the potential

multimodality of the low back vowels when conditioning on dialect areas it may result in more robust partitioning of acoustic space by social factors. By having two distinct central tendencies, listeners may be more likely to learn the patterns as belonging to separate generative models of talkers and being indicative of two production targets within the language. This perspective supports theoretical arguments that listeners track the distributions of the low back vowels within the community and may account for cases of near merger (Hay et al., 2009). Furthermore, the emergence of the low back vowels as two of the key indicators of dialect variation in this chapter supports current work that posits a structural relationship between the low vowels and regional shifts (Bigham, 2010; Kendall & Fridland, 2017). While this chapter doesn't evaluate the relationship between vowels, the findings here begin to elucidate the importance of the individual categories across regional vowel systems.

Conversely, the finding that Dialect is most informative of the cue distributions of /eɪ/ is surprising given the current literature on regional variation. Nevertheless, this finding suggests that regularity among talkers within dialect areas may occur in low-level variation that is not necessarily associated with salient shifts in the region. The noisy distributions from such a diverse dataset may of course provide a key context where such structure is likely to emerge across talkers despite such variable talkers. In addition, it's plausible that /eɪ/ may represent a category that is ultimately influential in the structural make-up of regional vowel spaces, but exactly how is unclear from these results alone. For example, the information gained for /eɪ/ may be indicative of the importance of peripherality in vowel shifts (Labov et al. 1972; Labov et al., 1991; Thomas 2001). The results in this chapter cannot directly speak to the exact cause or underlying source of the informativity and there is undoubtedly main potential explanations for /eɪ/. The implications of different types of structure across regional shifts warrants additional research to fully understand how variability (i.e., distributional properties) vary as a function of social and linguistic variables.

Critically, the results in this chapter illustrate regularity among talkers within geographical regions in their realizations of cue distributions across the vowel space. Such a finding is supported by two observations; 1) individual talkers' cue distributions are better estimated by the distributions of their region than the marginal distributions of American English or a single region (e.g., the West) and 2) talker divergence is lowest for categories that were most

conditioned by regional dialects. While individual vowel categories vary in terms of the magnitude of Talker divergence from dialect areas (Section 4.1.5 and 4.2.5), there is nevertheless an overall improvement across vowel categories. This result augments our current understanding of vowel shifts and lends support, at least broadly, to talkers' distributions mirroring their dialect areas despite some degree of heterogeneity among talkers within dialect areas. Given the continued debates around the role of individuals within communities in sociophonetics (see Chapter 2), these results suggest that while talkers may vary from one another in their central tendency, cohesion may be uncovered by examining the entirety of the distribution. However, analyses in this chapter still illustrated increased divergence of Talkers within specific dialect areas for vowel categories most saliently associated with the region. As such, research would benefit from continuing to identify mechanisms for uncovering these patterns of variability and the social sources of such divergences.

5.2 Contextualizing Socio-Indexical Structure: Perception

Using these simulations from production data, we can extrapolate hypotheses about listener behavior in speech categorization for further refinement and experimental testing (see Chapter 6 as an example). Informativity is conceptualized to estimate a lower bound of the degree of information to be gained from social factors in estimating cue distributions. In other words, we can use the analyses here as a proxy for listeners' prior experience to hypothesize a priori beliefs listeners may have about variability and the consequences they have on perceptual learning and generalization. As such, this section outlines how the various analytic scopes in this chapter depicts different a priori beliefs about socially constrained variation and predict variable listener behaviors.

The results presented in this chapter present a complicated picture for how listeners may approach the task of tracking talker variation. The results of this chapter align with some work in speaker perception, as indicated above. For example, there is alignment with listener perceptions of regional variation, such that the North and South appear as more divergent on average (Analysis 2, Section 4.2) which supports listeners' categorization of talkers into dialect areas (Clopper & Pisoni, 2004a-c, 2007) and subjective evaluations of regional variation (Preston,

1989, 1993). It remains unclear to, what extent listeners draw on such information for regulating adaptation and generalization behavior.

In the holistic contrast-based explanation, as described by Kleinschmidt (2019), listeners are predicted to draw on prior knowledge that vowels (broadly) vary substantially across talkers. Given that talkers vary substantially, previous experience with talkers may be less helpful to draw on, and all else being equal, listeners should adapt flexibility to the vocalic patterns of individual talkers across all vowel categories. Namely, all vowel categories should demonstrate the same degree of flexibility in learning with no asymmetrical patterns. In terms of cross-talker generalization, listeners should be more likely to generalize the patterns of an individual talker to other novel talkers provided that the talkers are perceived to be linguistically similar. As such, generalization is likely to be limited to same gender talkers due to the nature of gross cross-gender distinctions (Kleinschmidt, 2019) and acoustic similarity (Kraljic & Samuel, 2006; Liu & Holt, 2015; Reinisch et al., 2014; Xie & Myers, 2017). Overall, listeners should treat cross-talker variability of vowel patterns as one-size fits all: greater talker variability leading to more flexible adaptation and generalization behavior.

Alternatively, in all other cases, listeners may show asymmetrical learning or generalization of vowel categories based on whether the vowel category is conditioned on dialects (dialect-informative) or talkers (talker-informative). The dialect-informative categories can be further divided into dialect-agnostic and dialect-specific predictions. The dialect-agnostic perspective would suggest that listeners track variability across talkers and have generalized over experiences with talkers from various regions to form a causal link between variability and dialect background. This generalized causal model would assert that after continued exposure with a vowel category, for example /eɪ/, varying by talkers across different regional dialects, listeners would then learn that /eɪ/ variability is caused by speakers' regional background. Consequently, when exposed to a novel talker with an atypical /eɪ/ production, listeners would be likely to infer a priori that the variation is caused by the talker's regional background and learn the pattern as characteristic of the talker. Further, because variability is inferred to be driven by a social group, listeners should generalize the pattern to other talkers, assuming they infer they are sufficiently similar to the exposure talker.

On the other hand, the dialect-specific organization may provide more fine-grained causal links for listeners based on vowel-specific and dialect-specific combinations. This organization would suppose that as a listener encounters talkers from various backgrounds, they learn, for example, that /aɪ/ divergence is most likely caused by regional variation in the South. Similarly, they might learn that /a/ divergence is most likely caused by regional variation in the North. One prediction then is that listeners encountering talkers with atypical productions will show variable behavior for adaptation and generalization depending on how the talker is perceived relative to the regional pattern. For example, listeners may adapt quickly to atypical productions of /aɪ/ when the talker is inferred to be from the South, assuming all variation in /aɪ/ from Southern speakers is the same. Similarly, they will generalize the pattern but only to novel talkers perceived to be Southern. Alternatively, if listeners perceive the talker to be from the South and are faced with an atypical percept, they may be less likely to generalize due to their prior knowledge that talkers in the South are generally similar in their productions of /aɪ/. As such, they may infer the atypical production is idiosyncratic or otherwise incidentally caused and resist generalizing. This type of prediction is more aligned with expectations of sociophonetics, whereby listeners have some retention of the fine-grained variation of specific social groups.

Finally, the asymmetry between dialect-informative and talker-informative suggests that listeners may show asymmetrical behavior across vowel categories along these two dimensions. The Talker-informative hypothesis of variability would presume that the vowel category varies widely across talkers, but it is not conditioned by some larger socio-indexical group. (e.g., Dialect). Namely, listeners would have prior knowledge, for example, that /ʊ/ is variable across talkers and such variability is not caused by regional background. Consequently, when encountering a talker with an atypical /ʊ/, listeners would be likely to infer the pattern is characteristic of the talker and adapt quickly and robustly. Conversely, because /ʊ/ would not be inferred to be characteristic of a regional identity, listeners should learn in a talker-specific manner showing no generalization to novel talkers. In Chapter 6 I specifically test this hypothesis by examining perceptual learning across a dialect-informative category (/eɪ/) and a talker-informative category (/ʊ/); more specific predictions and further discussion are provided therein and in the final discussion chapter (Chapter 7).

Overall, different organizations of socio-indexical structure outline several possible listener behaviors depending on what listeners are tracking. The perceptual learning predictions outlined in this section are not exhaustive and there are potentially numerous additional listener behaviors that we might predict. However, they encompass a first instantiation of different listener behavior based on the causal inferences they may draw from tracking cross-talker variability informed specifically by the model outlined by Kleinschmidt (2019) and the analyses in this Chapter.

6 Conclusion

Overall, this chapter highlighted socio-indexical structure as an emergent property over the distributional properties of vowel categories holistically as it relates to different baseline experience. Over three analytic simulations, I identified an asymmetry between categories that can broadly be grouped into Dialect-informative and Talker-informative. Importantly, the dialect- informative categories are not necessarily those that are most saliently associated with the regional vowel shifts, but rather those that demonstrate consistent joint distributional cue properties across talkers within dialect areas. Categories like /eɪ/, for example, appear to be ranked highly in dialect informativity and demonstrate lower degrees of individual divergence from their regional groups. These patterns at the highest level suggest that there may be different types of socio-indexical variability and low-level variation is more likely to demonstrate coherence across talkers with a region. The results provide evidence that talkers generally align with their dialect areas in terms of the overall distributions of variability, as indicated by decreased divergence across categories when using talkers' dialect distributions as the reference distribution. Furthermore, the results of this chapter can be linked to some observed patterns of listeners' social perception, such as highly salient regional varieties. It remains to be seen whether such descriptions align with listener.

CHAPTER 5:
INTERNAL LINGUISTIC SPECIFICITY & SOCIO-INDEXICAL STRUCTURE

1 Introduction

In the previous chapter, I draw on an instantiation of the ideal adapter model, which posits that listeners accommodate acoustic overlap among vowel categories by tracking an *individual* vowel category's joint cue distributions (i.e., phonetic distributions of F1xF2) and the social factors that condition variability (Kleinschmidt 2019). Of course, an individual vowel category's phonetic cue distributions are part and parcel to how listeners identify a particular contrast and disambiguate it from other categories. However, such a model may assume that if an individual vowel category's distributions are not conditioned on social factors, then listeners are unlikely to leverage group level variation of the category for online processing.

As discussed previously (Chapter 2), sociophonetics has long since emphasized the internal principles of linguistic variation theorizing that vowels operate as a *system* of interrelated dependencies. Similarly, recent work examining structured variation has also described the considerable regularity of talkers as a function of phonetic dependencies (e.g., Chodroff & Wilson, 2022), such as individuals' mean F1 position correlating across vowels sharing the height articulatory dimension (e.g., Ahn & Chodroff, 2022; Ménard et al., 2008; Oushiro, 2019; Salesky et al., 2020; Schwartz & Lucie, 2019; Watt, 2000). Similarly, individual cue dimensions (as opposed to joint cue distributions) may show typological tendencies whereby cross-talker variation is more likely to occur along a specific cue dimension for specific categories. Thus, signaling the importance of socio-indexical conditioning along single cue dimensions for listeners' beliefs.

Thus, phonetic dependencies (i.e., relationships among vowels and cues) and individual cue dimensions may covary with social factors thereby facilitating online speech processing. And indeed, recent work has recognized that distributional properties of similar contrasts (Newark, 2001) and individual cue distributions (Idemaru & Holt, 2014; Reinisch et al., 2014) may constrain perceptual processes. While the ideal adapter model posed by some scholars (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015, Weatherholtz & Jaeger, 2016) largely

separates socio-indexical structure from other internal structure, there is of course many reasons to believe that these two facets interact and may be integral to listener beliefs. Indeed, much work has shown that listeners demonstrate both generalizability (i.e., across talkers and contrasts) while also showing a great deal of specificity in learning (i.e., cue distributions, talker-specific learning, etc.). The specificity of internally conditioned variation prompts more granular descriptions of socio-indexical structure and the generation of predictions for listeners' perceptual categorization and adaptation behavior.

To address this gap, in this chapter I will focus on describing specificity in two ways: (1) phonetic dependencies among vowels operationalized as acoustic overlap between vowel pairs and (2) cue distributions' shapes and parameters. Following the previous chapter, I examine different analytic methods for operationalizing socio-indexical groups over the frequency distributions of F1 and F2. In Part 1 (Section 2), I specifically examine how acoustic overlap between categories may decrease as a function of socio-indexical structure (e.g., dialects and talkers) and articulatory dimensions (e.g., front/back). In Part 2 (Section 3), I examine the distributional make-up of individual vowel categories to begin identifying specific dimensions of variability. I will review literature that grounds the analyses in-turn for each part (Section 2.1 and 3.1 respectively) followed by a discussion (Section 2.4 and 3.5 respectively) and conclusion.

2 Part 1: Acoustic Overlap Across Vowel Pairs & Socio-Indexical Factors

2.1 Introduction

Necessarily, listeners are disambiguating the speech signal to uncover a talker's intended linguistic message. The acoustic overlap among vowel categories (i.e., distributions in F1xF2 space) has long been established as a puzzle in speech perception since Peterson and Barney's (1952) foundational work describing American English vowels (and subsequent replications, e.g., Clopper et al., 2005; Hillenbrand et al., 1995; Hagiwara, 1997). The high degree of overlap among different vowel categories highlights the invariance problem among acoustic cues, that is: how do listeners accurately perceive speech with such variable realizations? The issue is further supported by perceptual categorization errors demonstrating that vowel categories are subject to a high degree of misidentification among listeners of English. The largest error rates in misidentification have been shown to occur in areas of the vowel space with higher

concentrations of neighboring vowels where acoustic overlap is greatest (e.g., mid-lax vowels; Peterson & Barney, 1952; Hillenbrand et al., 1995). On the other hand, categories with less overlap demonstrate decreased misidentification, such as /i/ and /a/ (Peterson & Barney, 1952). Discrimination of vowel categories may be improved when additional dimensions, such as vowel length, are considered (Ainsworth, 1972; Bennett, 1968; Labov & Baranowski, 2006; Fridland et al., 2014; Hillenbrand et al., 1995), which may aid in contrast maintenance over time (Labov & Baranowski, 2006; Pierrehumbert, 2001). Recent arguments about socio-indexical structure in perception suggest that socio-indexical information similarly aids in category discrimination and reduction in acoustic overlap.

Indeed, much of the acoustic overlap among category pairs is correlated with socio-indexical factors such as gender and dialect (Clopper et al., 2005; Hagiwara, 1997; Hillenbrand et al., 1995; Peterson & Barney, 1952). One criticism of Peterson and Barney's (1952) work was that regional variation was not well accounted for as a source of increased acoustic overlap, which subsequently sparked replications considering dialect variation more explicitly (Clopper et al., 2005; Hillenbrand et al., 1995; Hagiwara 1997). This work found that in fact some of the acoustic overlap among talkers was significantly reduced by accounting for dialect in addition to the effect of gender. Work in sociophonetics has continued to underscore this observation with proposed (and validated) methods for quantifying vocalic overlap (e.g., Hall-Lew, 2010). Thus, it's not unreasonable to assume that talker identity and socio-indexical groups may serve as an additional dimension by which listeners are able to disambiguate vowel categories, akin to duration.

Additionally, listeners may exploit the dependencies between vowel category distributions to resolve ambiguity of atypical productions of a given category (e.g., atypical /æ/ backing) by knowing the relative boundaries and the likelihood of encroaching on another category in phonetic space (e.g., /a/). Category pairs that tend to maintain higher degrees of separation across talkers may suggest that listeners show attenuated learning of a pattern that effectively increases the distributional overlap between otherwise separated categories due to its atypicality. Further, listeners may demonstrate a priori more robust categorization, showing greater certainty about category boundaries where the acoustic overlap is more minimal. Some work has suggested that listeners are more consistent in categorization for categories with

minimal overlap (Newman et al., 2001; Clayards et al., 2008), categorization boundaries further away from the vowels center of mass for more variable vowels (Chao et al., 2019), and adaptation is attenuated for distributions with more considerable overlap of the acoustic cue(s) (Drouin et al., 2016). Using experience with socio-indexical groups, listeners may further maintain category resolution by tracking the relative boundaries of vowel category pairs across different socio-indexical factors as overlap decreases as a function of, for example, regional background. Such an explanation may be elucidated by prior work in sociophonetics that has pursued how the relationships between vowels is socially structured and internally governed.

In particular, sociophonetic literature suggests that the maintenance of acoustic separation among certain vowel pairs is a functional outcome, if not mechanism, of vowel shifts. Principles of maximal dispersion, characterized as an effort to preserve contrasts and maintain and restore symmetry, have been described as a mechanism and/or outcome of vowel change. Maximal dispersion is reflected in the hierarchical (re)organization of phones within chain shifts along articulatory dimensions (e.g., frontness or height; Labov, 1994, 2001; Labov & Baranowski, 2006; Martinet, 1955). Labov (1994: 580–588) argues maximal dispersion is an automatic process of probabilistic matching between categories' distributional properties. Labov and Baranowski (2006: 223) illustrate this mechanism as follows:

> "Outliers from the normal distribution of realizations of a phoneme that overlap the normal distribution of a neighboring phoneme are less likely to be understood as being tokens of the intended phoneme, and are thus less likely to participate in the calculation of the mean target of that phoneme by the language learner. But when the neighboring phoneme has shifted away, increasing the margin of security, the same outlier is more likely to be recognized as a member of its intended phoneme, and thereby shift the calculation of the mean in its direction."

Accordingly, the boundaries of category distributions may contribute to how likely individuals are to assign individual experiences to a single category and in-turn update their representations. In cases where one phoneme has shifted and the overlap between the category distributions decreases, individuals may in turn demonstrate a greater likelihood of adapting to outliers and

updating categories. Namely, effects of maximal dispersion may effectively constrain distributional learning and phonetic retuning of an individual category.

Certain vocalic relationships in American English have specifically been posited to provide stability in the vowel space thereby maintaining principles of maximal dispersion. Specifically, Labov (2001) posits /æ/ is a critical category that may act as an anchoring category for other vowels in the system. One critical area where this process has been explored in production is the relationship between /æ/ and the low-back vowels. Several scholars have posited a structural relationship such that /æ/ retraction initiated realignment of the low back vowels (Bigham, 2010; D'Onofrio et al., 2016; Hall-Lew, 2013; Kendall & Fridland, 2017; Labov et al., 2006). As evidence, Bigham (2010) demonstrates that talkers who retracted /æ/ also merged low-back vowels, while the less retracted /æ/ talkers maintained unmerged, or only occasionally merged, low-back vowels.

The correlation of /æ/ and the low-back vowels has been hypothesized to emerge from a structural relationship of /æ/ and /a/, whereby movement in /æ/ triggered realignment of /a/ across several vowel shifts. In the case of the NCS, scholars have suggested that the fronting and raising of /æ/ triggered realignment and fronting of /a/ as the vowels move to fill empty gaps in the system, to fulfill principles of maximal dispersion of these two categories (Bigham, 2010; Thomas, 2011). Scholars have suggested the retraction of /æ/ in the LBMS triggered backing and raising of /a/, resulting in merger of /a/ and /ɔ/ (Bigham, 2010; Kendall & Fridland, 2017). Kendall and Fridland (2017) find evidence of this structural relationship such that the distance between /æ/ and /a/ categories is regular across dialect areas, despite the average positions of each vowel category differentiating regional systems. Further, listeners' degree of low-back merger in their own speech predicted their categorization boundaries along an /æ/-/a/ continuum. While this dissertation does not speak to the mechanistic accounts of maximal dispersion and vowel shifts, the static outcomes and distributional properties of the vowel systems may indeed provide listeners with stability from which to predict other vowel category distributions. As such, I examine the relationship of /æ/-/a/ in the analyses below as a specific example of such stability.

Of course, the reality is that vowel mergers still occur despite principles of maximal dispersion, and as such have received considerable attention in production and perception (Babel et al., 2013; DeCamp 1953; Hay et al., 2006; Hay et al., 2009; Labov et al., 1991; Trudgill, 1986;

Warren et al., 2007). Vowel mergers demonstrate unique patterns, with individuals' demonstrating different talker-listener behaviors whereby some people exhibit: 1) distinction in production and perception; 2) no distinction in production and perception; 3) distinction in production but lack of perceptual distinction; and 4) perceptual distinction but lack of distinction in production (DeCamp, 1953; Labov et al., 1991).

The asymmetry in perception and production (3 and 4) suggests that language users may be tracking the distributions of the categories in their communities, with differential experience with the merged and unmerged systems (Warren et al., 2007). Indeed, there is evidence that productions of merged /a/ and /ɔ/ vary depending on the talker, the community, and the overall degree or completeness of the merger (Herold, 1990). This suggests that the relationship between /a/ and /ɔ/ similarly demonstrates a high degree of socio-indexical conditioning which listeners attend to despite merged vowels being less socially salient (Labov, 2001). In addition, people who are merged in their dialect may become unmerged while accommodating to speech of unmerged talkers, with mediation from social cues providing expectations about the talker's merged status (Babel et al., 2013; Hay et al., 2006). Importantly for this dissertation, it suggests that people track the distributional properties of both categories to some extent, despite the potential for merged categories to not directly aid in disambiguation of contrasts, including for talkers who do not maintain the contrast themselves. Tracking this relationship thus may provide listeners the ability to make predictions about another category's distributions, provide the boundaries between the contrasts, and potentially aid social categorization of the talker.

Overall, socio-indexical structure among vocalic relationships may emerge in various ways including: 1) reduction in acoustic overlap of categories as a result of talker-specific or group conditioned distributional properties; 2) increase in overlap, particularly in cases of merger; 3) systematic separation of vocalic distributions among vowel pairs applied uniformly across talkers and groups. To rephrase, socio-indexical factors condition distributional properties of vowel pairs, and by tracking socio-indexical factors acoustic overlap may be systematically reduced or augmented (e.g., merged vowels). In addition, there may be internal principles that restrict the degree of overlap among certain vowel pairs, regardless of socio-indexical background (e.g., /æ/-/a/). Consequently, listeners may learn the boundaries of category membership. In cases where overlap is high, the boundaries between categories may become

fuzzier and increase misidentification and listeners may tolerate more ambiguity from categories that have higher degrees of overlap. On the other hand, when overlap is lowest, categories are less fuzzy and listeners may remain more rigid in category boundaries. In the latter case, listeners may extract stability from the vowel system which may effectively constrain adaptation and/or generalization of atypical forms that encroach on such a boundary. On the other hand, in the former, listeners may benefit from tracking socio-indexical patterns to better predict the category boundaries and disambiguate vowels across talkers effectively reducing the fuzzy boundaries.

The theoretical background outlined above motivates the need to examine relationships among vocalic contrasts as a source of socio-indexical structure, which may elucidate listener behaviors and allow for a more comprehensive integration of socio-indexical structure in inferential models of speech processing. In the following sections I specifically focus on how socio-indexical structure emerges through acoustic overlap across vowel pairs. The analyses follow a parallel logic of the previous chapter, in which I examine outcomes of statistical models across several socio-indexical grouping factors, examining both dialectal and talker-specific variation. Following the previous chapter, the different analyses are meant to provide descriptive accounts of variation across the vowel space, and as such I will not focus on significance testing between different socio-indexical granularities and leave this to future work. The discussion will focus on specific categories that have been shown to have higher rates of misidentification, the mid-vowels, to examine the extent to which socio-indexical structure reduces overlap among competing categories. I additionally discuss the relationship between /æ/ and /a/, as previous work has already highlighted and important relationship among these two categories. Within these analyses, I highlight the role of articulatory dimensions in the separation of vowel relationships, specifically focusing on the front/back dimension. Following the results, I will discuss implications for inferential speech processing providing some testable hypotheses that arise from these analyses.

2.2 Methods

2.2.1 Pillai

To quantify overlap across vowel pairs, I use the Pillai-Bartlett statistic (henceforth Pillai score), an output of the Multivariate Analysis of Variance (MANOVA) which indicates the

proportion of variance that can be predicted by a given factor, in this case vowel category (with two levels). A higher Pillai score indicates greater separation between the two categories (i.e., less overlap) and a lower Pillai score indicates more overlap between the categories. There are several other measures used to evaluate acoustic overlap across categories, but Pillai has been shown to perform best (Hall-Lew, 2010; Kelley & Tucker, 2020) and is used frequently in sociophonetics studies. Pillai scores are highly correlated with the more traditional R^2 to indicate variance explained by the independent variable. In this case, the dependent variable is the multivariate cues, F1xF2, and the independent variable is the contrast between two vowel category pairs (e.g., /æ/ and /a/).

2.2.2 Defining Groups

As noted throughout this dissertation, a central interest in describing socio-indexical structure is considering how individuals fit within their broader group patterns, and how we conceptualize the group structure more generally. In Chapter 4, using KL divergence, socio-indexical organization was flat, whereby tokens across talkers all add to cumulative distributional properties without parametric characterizations of individual talkers' distributions within their dialect group. I will follow the same conceptualization here, where the organization is flat across the overall dataset and dialect areas. Contrasts are often treated as properties of individual talkers, not necessarily properties of larger group structures, and as such Pillai is typically calculated over individual talkers. By applying Pillai scores to broader socio-indexical factors, this adds both a methodological and theoretical component in describing how socio-indexical factors contribute to acoustic overlap.

I extend the logic of KL divergence in the previous chapter here, testing whether group level distributions of vowel cues across pairs is increased/reduced with a flat dialectal level grouping structure, compared to the broader dataset. I consider all tokens across all talkers as the overall distribution, similar to the marginal distribution in the previous chapter, but will refer to it as 'All data' in figures. Thus, the datasets distributional overlap across vowel pairs serves as a 'floor' from which higher-order indexical factors may contribute to changes in the degree of acoustic overlap. The higher-order indexical factors are organized accordingly with Chapter 4, with the Dialect factor representing an aggregate measure of the individual dialect levels (i.e., the

average) and ‘Talker’ as an aggregate measure of individual talkers. The Dialect factor assumes the same flat structure to KL divergence, where individual overlap properties across vowel pairs are calculated over all tokens across all talkers for a given dialect level. The Talker factor is calculated over vowel pairs on an individual talker basis. Replicating the same pattern in Chapter 4, I provide both the dialect-agnostic perspective (i.e., Dialect) alongside the dialect-specific perspective (i.e., individual dialect areas). As such, some figures below will depict overall averages (e.g., average Dialect Pillai score for /æ/-/a/), while others will represent factor levels (e.g., Pillai score for /æ/-/a/ in the South), as indicated by section headings and figure captions.

2.3 Analyses

2.3.1 American English (All Data)

To begin this section, a brief overview of the type of variability within categories we see across the dataset (tokens and talkers) is shown below in Figure 5.1. Figure 5.1 shows the vowel space across all of the data (from Chapter 3) with ellipses representing the 95% CI under a multivariate normal distribution for each vowel category. In line with previous work (e.g., Peterson & Barney, 1952; Hillenbrand et al., 1995), this figure illustrates the high degree of variability across vowel categories at a larger scale and demonstrates a key problem for perception with acoustic cues to vowel identity overlapping. The large areas of overlap are, at least visually, not limited to neighboring vowels or those that share the same articulatory dimensions. Rather, categories across back and front as well as high and low demonstrate considerable overlap when examined across the aggregate data. Additionally, and unsurprisingly, due to the variety of speech styles, number of talkers, and variable lexical items, the degree of overlap is more extensive than previous work examining lab speech has depicted (e.g., Clopper et al., 2005; Hillenbrand et al., 1995; Peterson & Barney, 1952).

The size and respective degrees of overlap depicted by the ellipses in Figure 5.1 correspond with the Pillai scores quantifying the degree of overlap among different vowel pairs across the vowel space, reported in Table 5.1. As evident from Table 5.1, Pillai scores range from 0.14, indicating near complete overlapping distributions (e.g., /ʌ/ and /o/), to 0.88, indicating little overlap in the distributions (/ɔ/ + /i/). Category pairs that demonstrate the greatest degree of overlap are neighboring vowel categories (e.g., /i/ +/ɪ/ Pillai = 0.34) and category pairs that are

concentrated at the middle of the vowel space (e.g., /ɔ/ + /ʌ/). The acoustic overlap among vowel category pairs is greatest along distinct articulatory dimensions, such that vowel pairs spanning either high-low (e.g., /u/ + /a/ Pillai = 0.80) or front-back (e.g., /ɔ/ + /eɪ/ Pillai = 0.85) dimensions have greater acoustic difference compared to pairs that share an articulatory dimension (e.g., front vowel pairs /i/ + /eɪ/ = 0.34). Unsurprisingly, Pillai scores are lowest when pairs are differentiated by more distinctive articulatory features, demonstrating less acoustic overlap and greater separation.

Given the extensive overlap of vowel pairs, I turn to assess whether socio-indexical factors structure distributional properties by altering the amount of overlap between pairs of categories with a high degree of acoustic overlap. To do so, I turn to examine how conditioning on socio-indexical factors influences Pillai scores. From this point forward, I will focus only on specific vowel pairs that may provide higher degrees of challenge for listeners, encompassing pairs along the middle portion of the vowel space (e.g., /eɪ/, /ɛ/, /ʊ/, /o/), and highlight the degree to which attending to socio-indexical factors would decrease the degree of overlap. In addition, I will also examine the relationship between /æ/ and /a/, due to the suggested special relational status. Following a similar logical progression as Chapter 4, I will examine the dialect-agnostic perspective (i.e., the aggregate across dialect levels) in Section 2.3.2, followed by an examination of dialect-specific trends in 2.3.3. Finally, in Section 2.3.4 I examine how the factor of Talker structures the distributional properties of vowel pairs.

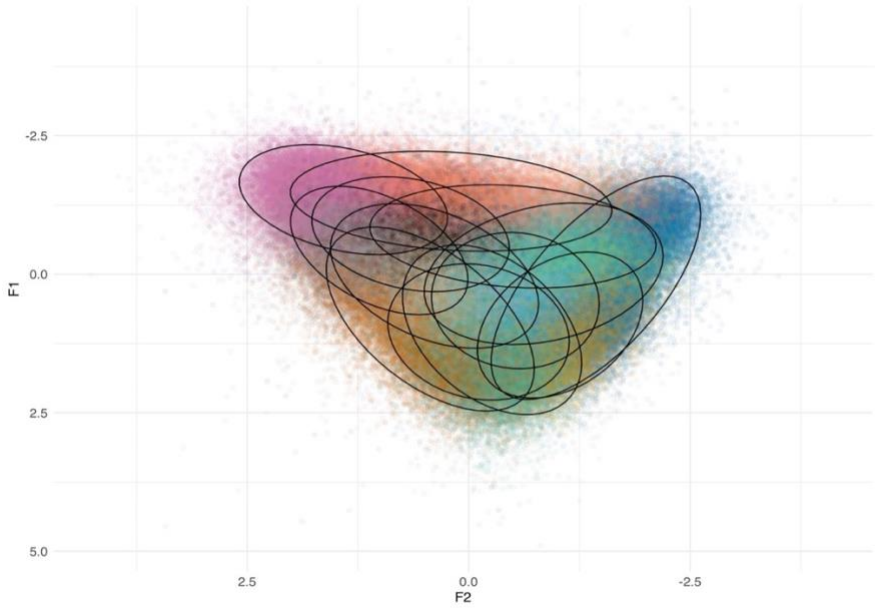

Figure 5.1 Vowel space across all American English vowels, talkers, and tokens from the dataset in Chapter 3. Ellipses represent the 95% CI around the mean assuming a multivariate normal distribution.

Table 5.1 Pillai scores across vowel pairs and all data (not sub-set or conditioned on social factors).

	a	æ	aɪ	eɪ	ɛ	i	o	ɔ	u	ʊ	ɪ
ʌ	0.29	0.53	0.23	0.72	0.39	0.80	0.14	0.43	0.69	0.30	0.64
a		0.65	0.38	0.82	0.61	0.86	0.36	0.18	0.80	0.56	0.8
æ	0.65		0.30	0.41	0.26	0.70	0.59	0.72	0.80	0.61	0.54
aɪ	0.38	0.30		0.64	0.35	0.78	0.45	0.57	0.72	0.43	0.62
eɪ	0.82	0.41	0.64		0.25	0.39	0.72	0.85	0.59	0.56	0.22
ɛ	0.61	0.26	0.35	0.25		0.59	0.48	0.85	0.59	0.56	0.25
i	0.86	0.70	0.78	0.39	0.59		0.80	0.88	0.41	0.48	0.34
o	0.36	0.59	0.45	0.72	0.48	0.80		0.27	0.62	0.18	0.64
ɔ	0.18	0.72	0.57	0.85	0.70	0.88	0.27		0.73	0.45	0.81
u	0.80	0.74	0.72	0.59	0.56	0.41	0.62	0.73		0.24	0.35
ʊ	0.56	0.61	0.43	0.56	0.32	0.48	0.18	0.45	0.24		0.33
ɪ	0.80	0.54	0.62	0.22	0.25	0.34	0.65	0.81	0.35	0.33	

2.3.2 Dialect-Agnostic

Figure 5.2 below shows the Pillai scores of vowel pairs, with the blue circles representing the Pillai scores for each vowel pair given the entire dataset (the values in Table 5.1 above). The orange squares represent the Dialect factor, as an average Pillai score for each vowel pair across dialect levels, which are further indicated in Table 5.2 below. Pillai scores were calculated for each vowel pair (e.g., /a/ and /ɔ/) conditioned on each dialect area (e.g., South) and then averaged for each vowel pair across dialects (e.g., South + Midland + etc. for /a/ and /ɔ/, divided by the total number of dialect areas, 7). The entire dataset distribution (blue circles) represents a single value calculated for each vowel pair, as they were calculated across all data, replicated from Table 5.1 above.

Figure 5.2 shows that for many of the vowel pairs, the Dialect factor shows minor shifts across some vowel pairs while others remain stable. In line with the overall section above, the Dialect factor demonstrates a similar pattern of lower acoustic overlap among vowel pairs that differ along articulatory dimensions (e.g., high-low or front-back pairs). More specifically, category pairs that show the least acoustic overlap (e.g., /i/ + /ɔ/ Pillai = 0.90) in the overall data distributions maintain similar degrees of separation across dialects and align with the separation of articulatory dimensions. This is unsurprising given that Pillai scores have an upper-bound of 1 and any increase to already maximally distinct categories would likely be minimal. Further, we might expect that categories that are more articulatory distinct would be less likely to overlap.

Turning to the relationship of /æ/-/a/, we see that the acoustic overlap across the overall distributions and conditioned on Dialect is minimally different (Pillai = 0.65 and 0.69 respectively). Thus, the acoustic separation between /æ/-/a/ is relatively similar when conditioning on Dialect compared to the overall distribution. Similarly, /a/-/ɔ/ shows highly overlapping distributions across the overall dataset (Pillai = 0.18) and is minimally improved by conditioning on Dialect (Pillai = 0.23). The high overlap of /a/-/ɔ/ not being decreased by the Dialect factor is not surprising, given that many of the Dialects represented in the dataset merge these two categories. As such, it's possible that any differences in overlap among /a/-/ɔ/ and /æ/-/a/ across Dialects is not apparent in the aggregate and will only be apparent when examining dialect-specific patterns (see Section 2.3.3). Overall, the pattern of /æ/-/a/ may reflect the overall

separation of categories along distinct articulatory dimensions (i.e., front vs. back), in-line with other similar vowel pairings already observed.

Table 5.2 Mean Pillai scores across vowel pairs calculated over dialect levels, representing the dialect-agnostic perspective and the Dialect factor.

	A	æ	aɪ	eɪ	ɛ	i	o	ɔ	u	ʊ	ɪ
ʌ	0.30	0.55	0.25	0.77	0.42	0.84	0.20	0.44	0.72	0.32	0.69
a		0.69	0.42	0.86	0.65	0.90	0.39	0.23	0.74	0.59	0.83
æ	0.69		0.31	0.48	0.26	0.74	0.64	0.75	0.76	0.61	0.56
aɪ	0.42	0.31		0.69	0.34	0.81	0.51	0.6	0.74	0.45	0.62
eɪ	0.86	0.48	0.69		0.36	0.43	0.79	0.88	0.62	0.59	0.25
ɛ	0.65	0.26	0.34	0.36		0.66	0.54	0.71	0.61	0.36	0.3
i	0.90	0.74	0.81	0.43	0.66		0.85	0.90	0.49	0.57	0.41
o	0.39	0.64	0.51	0.79	0.54	0.85		0.24	0.65	0.19	0.7
ɔ	0.23	0.75	0.60	0.88	0.71	0.90	0.24		0.75	0.44	0.83
u	0.83	0.76	0.74	0.62	0.61	0.49	0.65	0.75		0.31	0.4
ʊ	0.59	0.61	0.45	0.59	0.36	0.57	0.19	0.44	0.31		0.39
ɪ	0.83	0.56	0.62	0.25	0.30	0.41	0.70	0.83	0.4	0.39	

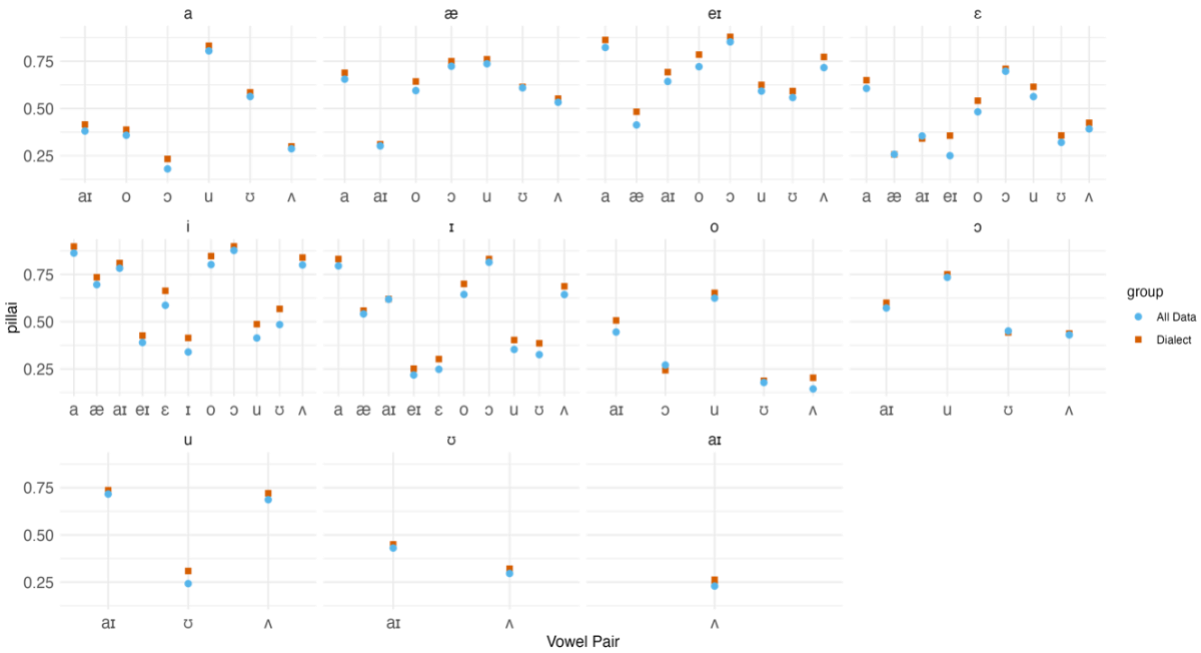

Figure 5.2 Pillai score across vowel pairs when calculated across all data (blue circles) and the average Pillai scores calculated across dialect levels (orange square)

2.3.2.1 Middle of the Vowel Space

For category pairs where there is greater acoustic overlap, we may hypothesize that there should be a change in Pillai scores, whereby Dialect factor shows higher Pillai scores, if Dialect aids in separating the acoustic overlap among pairs. At this point, I will examine the middle of the vowel space to examine an area for which there is a high degree of overlap among vowels, for which I've included mid front and back vowels, and /ʌ/ (as a central vowel) and /ʊ/. Figure 5.3 shows a decrease in acoustic overlap among pairs with front mid-vowels when conditioned on Dialect, varying in magnitude. Further, the front vowels /ɛ/ and /eɪ/ predominately follow the observation above, whereby acoustic overlap is least among category pairs that differ in an articulatory dimension, such as with back vowel pairs (in-line with the 'All Data' case). One exception is that /ɛ/ shows high overlap with /ɔ/ across All Data (Pillai = 0.34), which is not attenuated by conditioning on Dialect (Pillai = 0.43). Further, Dialect conditioned back vowel pairs show comparable degrees of acoustic overlap with the overall dataset distributions, showing no reduction in acoustic overlap.

In terms of front vowel pairs, we see greater separation when conditioned on Dialect. For example, /ɛ/-/ɪ/ overlap slightly decreases when conditioned on Dialect (Pillai = 0.30) compared

to All data (Pillai = 0.25), but only minimally. The greatest change in Pillai can be observed between /eɪ/ and /ɛ/, where overlap is high ('All Data' Pillai = 0.25) and shows a moderate decrease when conditioned on Dialect (Pillai = 0.36). The back vowel pairs, /ɔ/, /ʊ/, and /o/ on average show higher overlap with one another than front vowel pairings and overlap is not decreased when conditioned on Dialect. Overall, these results validate observations from Chapter 4, where /eɪ/ distributions are conditioned by Dialect and functionally decrease overlap (albeit only slightly) among neighboring vowel pairs. In contrast to Chapter 4, these data illustrate that /ɛ/ distributions may also be conditioned on Dialect, given attenuation of overlap with neighboring vowels across Dialects. Overall acoustic overlap is slightly reduced when conditioning on Dialect compared to the overall data set for the mid front vowels. This supports the hypothesis that socio-indexical factors may act as an added dimension to reduce category overlap across vowel pairs. However, the same is not true for mid-back vowel pairs, where acoustic overlap is maintained. As such, listeners may benefit from tracking /eɪ/ and /ɛ/ across dialect areas to improve disambiguation of other vowel categories, I will return to this in the interim discussion (Section 2.4).

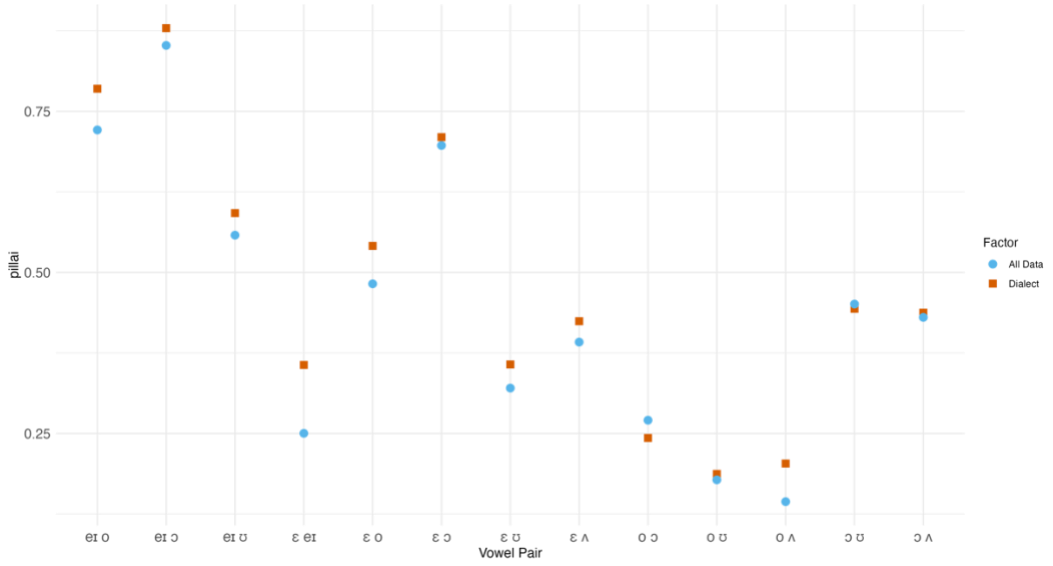

Figure 5.3 Pillai score across vowel pairs when calculated across all data (blue circles) and the average Pillai scores calculated across dialect levels (orange square), subsection of vowel pairs in the middle of the vowel space.

To summarize, there is some indication that the mid-front vowels show some decrease in acoustic overlap across neighboring vowel pairs when conditioned on Dialect while the back-vowels don't show the same attenuation of acoustic overlap. This suggests that listeners may be more likely to attend to talker-specific distributions for the back vowels, as Dialect groups alone is not enough for adequate disambiguation. While, on the other hand, attending to Dialect distributions may aid in disambiguating pairs of front vowels. Finally, there is evidence that acoustic separability of some vowel pairings is a function of differences in articulatory dimensions—that is the acoustic cues align with a general separation of front/back and high/low and such distinction is generally maintained across the overall distributions and when conditioned on Dialect. This pattern suggests that listeners may have expectations about the boundaries of variability, such that vowel pairings are less likely to overlap when they are cross articulatory dimensions (see Section 2.4 for additional discussion).

2.3.3 Dialect-Specific

In this section, I consider how a more hierarchical relationship among dialects and talkers provides more substantive structure among vocalic relationships. The first area of this hierarchical relationship I will explore is how dialect-specific patterns provide information over an aggregate dialect-agnostic perspective (as described in Chapter 4). I hypothesize that an average overlap statistic across dialects may capture the structure necessary for attenuating acoustic overlap due to the variable patterns across dialect areas. In particular, acoustic overlap may be attenuated or augmented for different dialect areas for the same vowel category pairs, such as the low-back vowels. Thus, the overall dialect average (dialect-agnostic) may effectively mask dialect-specific patterns.

Figure 5.4 below shows Pillai scores for each dialect level, across all seven dialect areas. Figure 5.4 largely aligns with the patterns observed in the previous section, with more overlap among vowel pairs sharing the same articulatory dimension than those separated by articulatory dimension for individual dialect areas. We also see that overall /æ/-/a/ overlap remains relatively constant across the dialect areas, despite variable positions for the two vowel categories. That is, regardless of where /æ/ and /a/ are positioned in the vowel space across dialect areas, the degree of acoustic overlap between the two vowels remains constant. In contrast, the low-back vowels

demonstrate variable degrees of overlap by dialect area, in-line with variable participation of the low-back merger across dialect areas. This finding generally supports the argument that the relationship between /æ/-/a/ is relatively stable across dialect areas, regardless of the relative shifts of the two vowels. In other words, even as /æ/ becomes more /a/ like acoustically (e.g., the LBMS), the distributions of /a/ and /æ/ maintain separation. Such a pattern suggests some stability in the vowel space which listeners may exploit to predict boundaries of variation across the two categories (see Section 2.4).

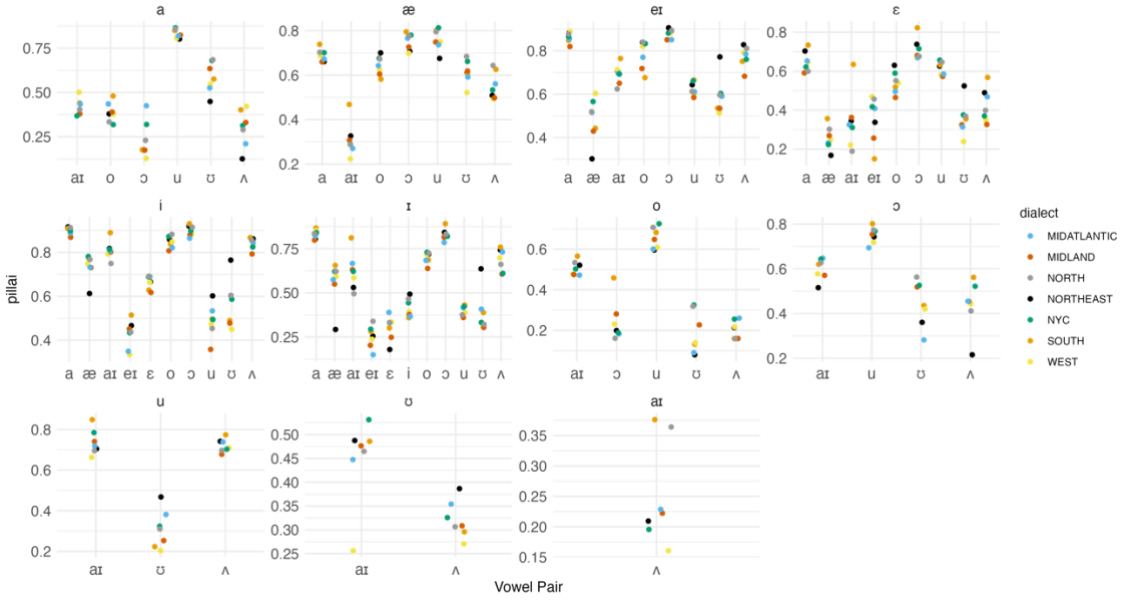

Figure 5.4 Individual Pillai score across vowel pairs when calculated across dialect levels, with individual points representing the dialect area.

2.3.3.1 Middle of the Vowel Space

In addition, there are a few interesting observations from the dialect-specific patterns when looking at the middle of the vowel space. In Figure 5.5, we see that there is variability among Pillai scores across dialects in terms of the degree of overlap among different vowel pairs. In some cases, dialect levels reduce the amount of overlap between pairs, as in /o/-/ɔ/ in the South (Pillai = 0.46) and in other cases there is an increase in the amount of overlap between categories, as is the case for /eɪ/-/ɛ/ in the South (Pillai = 0.15). Such a pattern largely aligns with expectations of the SVS, where spectral overlap is greater among the mid-front vowels compared

to other dialects as a critical facet of the SVS, and where /o/ fronting may decrease overlap with /ɔ/. The variable degrees of overlap among different pairs further illustrates that the acoustic overlap among back vowel pairs may be reduced or increased at the individual dialect level. Despite shifts in Pillai not being apparent at the dialect-agnostic level, individual regions provide varying degrees of attenuation or augmentation of overlap. The high degree of dialect-specific variability may suggest that listeners are more likely to attend to dialect-specific patterns or talker-specific patterns rather than relying on dialect-agnostic boundaries which may not provide enough input to effectively disambiguate category pairs.

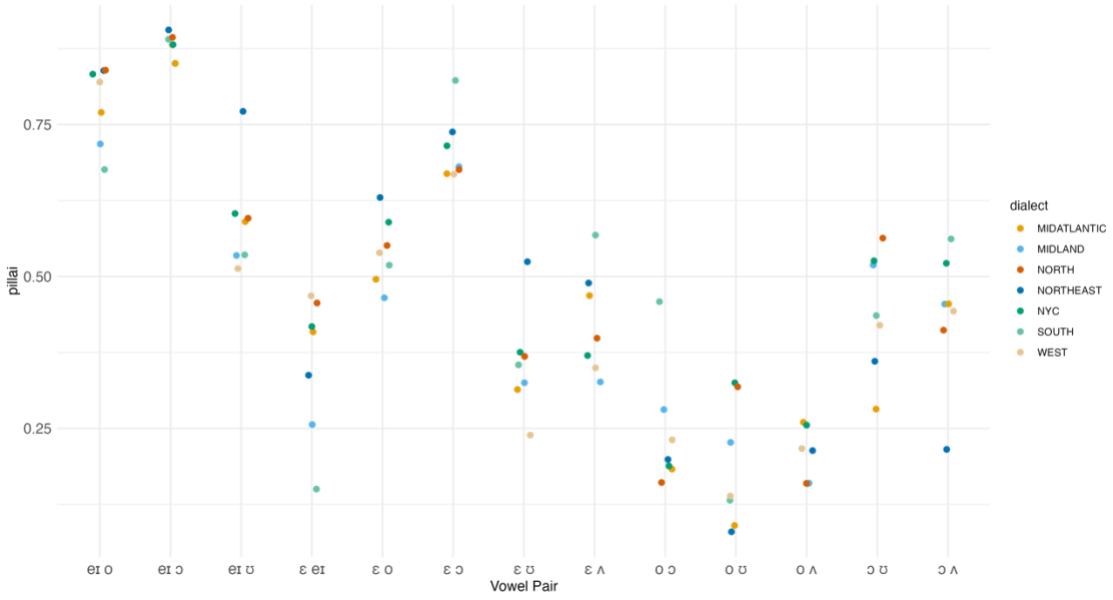

Figure 5.5 Individual Pillai score across vowel pairs when calculated across dialect levels, with individual points representing the dialect area, subsection of vowel pairs in the middle of the vowel space.

2.3.4 Talkers

Turning our attention to talker-specific patterns, this section examines Pillai scores calculated for each talker for each vowel pair. Figure 5.6 illustrates average Pillai scores conditioned on individual talkers (illustrated in grey squares), alongside the All data and Dialect factor analyses from the previous sections (again, blue circle and orange triangle respectively). Figure 5.6 illustrates that conditioning on talker (‘Talker’) decreases the distributional overlap of vowel category pairs overall. In-line with the previous section, there continues to be lower

degrees of overlap (i.e., greater separation) in pairs differing along articulatory dimensions, and these values generally remain the same regardless of conditioning factors. As depicted in Figure 5.7 and supported by Table 5.3, the spread of individual talkers' Pillai scores conforms to this pattern more broadly. Figure 5.7 illustrates the overall probability density distribution, a point representing the mean (along the bottom of each density plot), and the credible intervals. The credible intervals are analogous to frequentist confidence intervals but denote where points within a given interval (e.g., 95% CI) have a higher probability density than points outside of the interval (e.g., remaining 5%), estimating the relative bounds of the distribution.

For category pairs that span front/back dimensions there is generally decreased variability of Pillai scores across talkers, suggesting across talkers, regularity can be seen dividing the vowel space into subsystems defined by articulatory dimensions (i.e., front/back and high/low). Finally, examining the relationship between /æ/ and /a/, we find a similar degree of separability between these categories across individual dialect levels (Figure 5.5) as across talkers (Figure 5.7). While the /æ/-/a/ Pillai scores are not the highest degree of separability of all vowel pairs (here, /i/-/a/), they are greater than several other vowel pairs that share height dimension and only differ in the front/back dimension. Additionally, while there is some degree of individual variability in Pillai scores, they remain relatively low (Pillai = 0.12) and the values remain similar across all other grouping factors both on average, and across different levels of the 'Dialect' factor (i.e., individual dialect areas). This supports the suggested important status of /æ/-/a/ as providing some internal regularity to the vowel system and maintenance across vowel shifts, which is maintained from the overall distributions to individual talkers' distributions. Such a result suggests that listeners may have some expectations about the likelihood of /æ/ and the relative boundaries of /a/ across talkers.

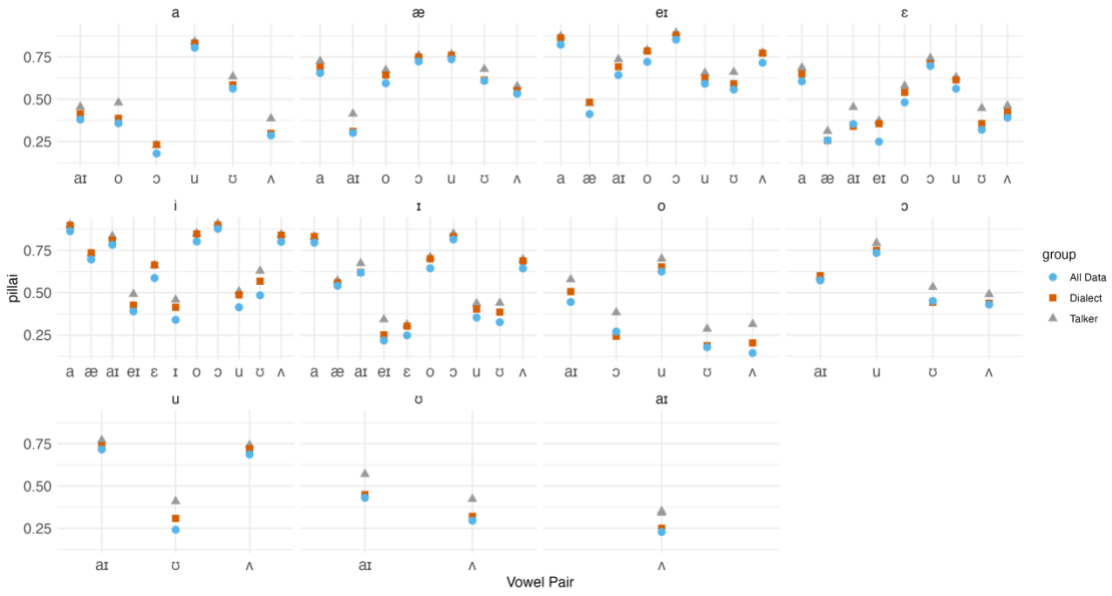

Figure 5.6 Pillai score across vowel pairs when calculated across all data (blue circles) and the average Pillai scores calculated across dialect levels (orange square), and Talkers (grey triangle)

Table 5.3 Mean Pillai scores across vowel pairs calculated over individual talkers, representing the Talker factor.

	a	æ	aɪ	eɪ	ɛ	i	o	ɔ	u	ʊ	ɪ
ʌ	0.39	0.58	0.34	0.78	0.46	0.84	0.31	0.49	0.74	0.42	0.7
a		0.73	0.46	0.87	0.69	0.9	0.48	0.23	0.84	0.63	0.83
æ	0.73		0.41	0.48	0.31	0.73	0.67	0.67	0.77	0.68	0.57
aɪ	0.46	0.41		0.74	0.45	0.83	0.58	0.59	0.77	0.57	0.67
eɪ	0.87	0.48	0.74		0.37	0.49	0.79	0.89	0.66	0.66	0.34
ɛ	0.69	0.31	0.45	0.37		0.66	0.58	0.74	0.63	0.45	0.31
i	0.9	0.73	0.83	0.49	0.66		0.85	0.91	0.51	0.63	0.46
o	0.48	0.67	0.58	0.79	0.58	0.85		0.38	0.7	0.29	0.71
ɔ	0.23	0.76	0.59	0.89	0.74	0.91	0.38		0.79	0.53	0.85
u	0.84	0.77	0.77	0.66	0.63	0.51	0.7	0.79		0.41	0.44
ʊ	0.63	0.68	0.77	0.66	0.45	0.63	0.29	0.53	0.41		0.44
ɪ	0.83	0.57	0.67	0.34	0.31	0.46	0.71	0.85	0.44	0.44	

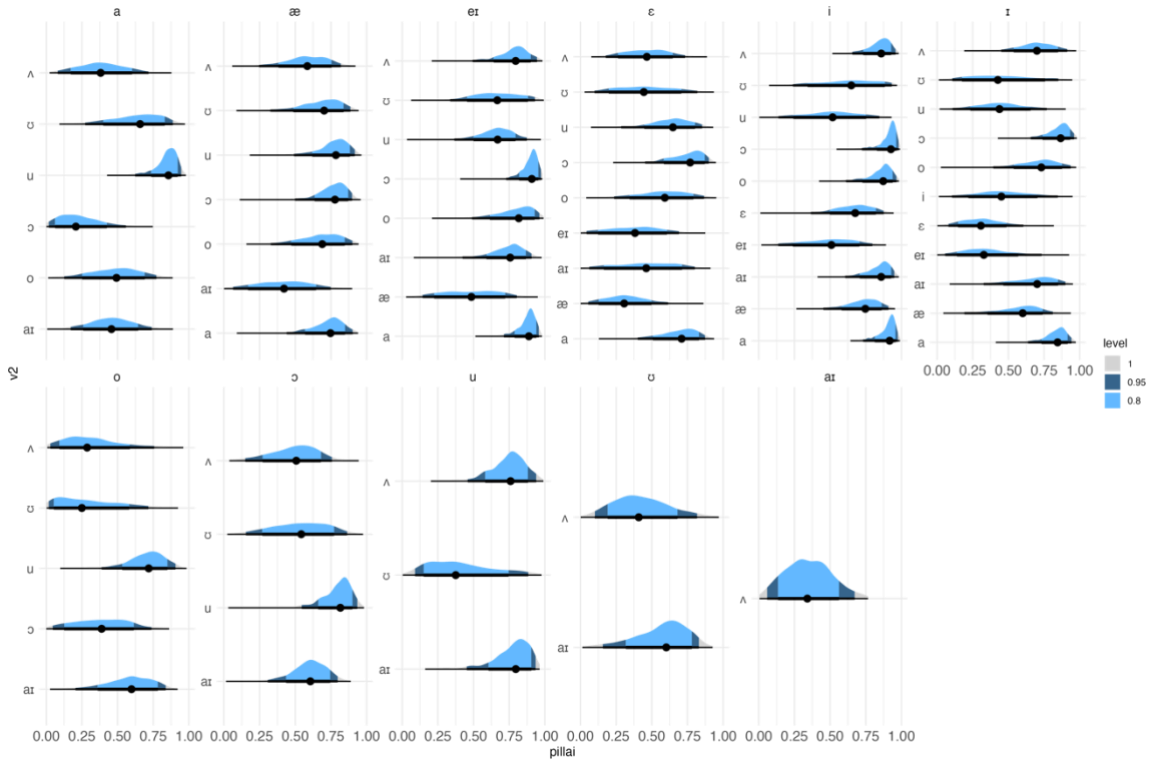

Figure 5.7 Half stat-eye density distributions of individual talkers' Pillai scores across vowel pairs. Color indicates confidence level, bright blue representing 80%, dark blue indicating the remainder to 95%, and grey 100%. Points along the x-axis represent the mean and whiskers represent the interquartile range.

2.3.4.1 Middle of the Vowel Space

In the middle of the vowel space, there appears to be minimal changes to Pillai scores across vowel pairings compared to the Dialect factor (as illustrated in Figure 5.6 above). For example, the vowel pair /eɪ/-/ɛ/ does not demonstrate a decrease in acoustic overlap when conditioned on Talker (0.37) compared to conditioning on Dialect (0.32) but is still decreased in comparison to the overall data distribution (i.e., All data). Similarly, vowel pairings such as /eɪ/-/o/ and /ɛ/-/o/ demonstrate some decrease in overlap properties when conditioned on 'Talker' (0.79 and 0.58 respectively) but remain on par with that of Dialect. On the other hand, Talker provides the greatest reduction in overlap among back vowel pairings, where conditioning on Dialect provided no change in relation to the overall distribution. For example, /ɔ/+/o/ overlap is slightly decreased when conditioned on 'Talker' (0.34) compared to the 'Dialect' (0.24) or 'All

data' (0.27) conditions. Figure 5.7 further illustrates that talkers are highly variable in terms of the degree of acoustic overlap across vowel pairings near the middle of the vowel space.

Overall, emergent structure among vowel pairs is illustrated through a reduction in acoustic overlap among the highly variable and overlapping distributions of the mid-vowel system, albeit not robustly. Conditioning on Talkers or Dialect groups provides similar reduction in acoustic overlap across vowel pairs containing front vowels. On the other hand, the acoustic overlap among pairs containing a back vowel demonstrates that Talker identity results in the greatest reduction of acoustic overlap. Similarly, there continues to be a high degree of phonological regularity in the system where separation of the mid vowels is greater for category pairs that span the front/back dimension which remains stable across socio-indexical factors. Finally, the results generally align with the suggestions that the relationship between /æ/ and /a/ tends to be highly structured across talkers and dialect areas, such that regardless of shifted positions, the degree of overlap among different vowel categories is generally preserved across various socio-indexical factors and levels. Below, I will discuss the implications of these findings for inferential speech processing, especially as it pertains to perceptual learning.

2.4 Interim Discussion

Overall, there is some support here for socio-indexical structure reducing acoustic overlap among mid vowels across socio-indexical factors, though only to a small degree. A reduction is observed among mid-vowel pairs that include a front vowel across both Talkers and Dialect factors. However, the back vowel pairs only show a reduction in acoustic overlap when conditioned on Talker and not Dialect. These results suggest that different vocalic relationships may be more sensitive to talker-specific and dialect-agnostic learning than others. Listeners may be more likely to draw on cross-talker expectations when resolving ambiguity for front vowel pairs and less likely to use expectations from other talkers when resolving ambiguity across back vowel pairs. Though, the degree of inter-dialect and inter-talker variability among Pillai scores suggests that listeners may generally not rely on cross-talker expectations to resolve ambiguity. That is, the high degree of instability of dialect information for vowels in the middle of the vowel space may lead listeners to rely less on cross-talker expectations of vowel category variability therein.

The fact that the mid-vowels demonstrate moderate acoustic overlap with one another that is variable in degree of socio-indexical factors reduction may lead to some hypotheses about perceptual learning. In particular, we might hypothesize that the boundaries between categories are fuzzy as a result of more extensive acoustic overlap compared to other vowel pairs in the system. As such, we might predict that listeners will be more tolerant of variants that encroach on the space of the neighboring vowel category a priori because these categories are highly variable across talkers. Subsequently, listeners rely less on their prior experience with social factors and cross-talker variation in disambiguating categories. As such, ambiguity is more likely to persist across talkers among certain vowel pairs, and/or disambiguation is resolved using other cue dimensions (e.g., duration, Wade, 2017), and/or higher-order levels such as syntactic (Clark, 2013) or semantic predictability (Borsky et al., 1998; Kalikow et al., 1977; McAuliffe 2015; Samuel, 1981).

Take for example the relationship between /eɪ/-/ɛ/, where there is considerable acoustic overlap that is only partially reduced by socio-indexical factors on average, and the degree is variable across talkers and dialects. Due to the large degree of overlap among pairs, it's possible that the boundaries of variation for /eɪ/ and /ɛ/ are much fuzzier, and listeners would have a higher degree of uncertainty around category membership and higher credibility may be given to both hypotheses about the category belonging to /eɪ/ or /ɛ/. This prediction aligns with prior work in distributional learning, where tokens are more readily assigned to distributions with greater variability where the likelihood of outliers in the category is higher, even when the item is more similar to the lower variance set (see Kapatsinski, 2018). Similarly, categorization functions for learned contrasts are steeper (i.e., greater certainty and more stability) for low variance compared to high variance exposure and greater overlap of categories (Clayards et al., 2008; Newman et al., 2001). Furthermore, since the variance and overlap of the categories is high, the reliability of the cues in distinguishing one contrast from the other is more likely minimal (Allen & Miller, 2004; Clayards et al., 2008; Newman et al., 2001; Kleinschmidt & Jaeger, 2015). As a result, listeners may be generally less likely to update their beliefs about the category and show minimal-no learning (Kleinschmidt & Jaeger, 2015).

On the other hand, and unsurprisingly, the phonological system provides a great deal of regularity in the system where overlap among vowel pairs is reduced when they vary along an

articulatory dimension (e.g., high/low, front/back). In addition, the results here support previous work that suggests the relationship between /æ/ and /a/ is structured across talkers and dialect areas, as is evident in an overall greater separability between the multivariate distributions. Regardless of the spectral position of /æ/ and /a/, the degree to which the acoustic distributions overlap is maintained across dialect areas and, for the most part, talkers. The internal category structure and covariation of vowel categories may further provide stability and input about the relative boundaries of variability across vowel pairs. As a result, such systematicity may further constrain perceptual learning.

The greater separation of acoustic cues across vowel pairs that differ in articulatory dimension may provide listeners with regularity across talkers, where the boundaries of variability between pairs is more distinct. We may predict that due to more limited cross-talker variation across these boundaries, listeners may have a harder time adapting to shifts towards categories that are otherwise stable across talkers. To illustrate this idea, take for instance the vowel pair /eɪ/ and /o/, which differ along front/back dimensions but share height dimensions. This pairing generally demonstrates greater separability in cue distributions (according to the analyses presented here). As such, listeners may show more difficulty in adapting to a pattern whereby /eɪ/ becomes more like /o/, or vice versa (particularly under complete remapping), because that would be both uncommon in experience and violates properties of the phonological system. If listeners learn this pattern, it may be limited to a talker-specific pattern and listeners may be unlikely to generalize the pattern to other talkers. This may be driven by several factors, including the possibilities outlined by principles of maximal dispersion, but also a result of listeners not inferring variability is caused by dialectal variation. Similarly, this may be true for categories like /æ/ and /a/ where the structural relationship is highly regular and structured across talkers and shifts. However, this pattern may be limited to cases where both distributional patterns are present, such that listeners receive input about the distributions of both categories, which may more strongly indicate a merger of the categories.

Ambiguity is thus maximized for category pairs where acoustic overlap is greatest and where conditioning on socio-indexical factors only minimally reduces the overlap. Whereas ambiguity is inherently lower when acoustic overlap between pairs is low, which may restrict the range of acceptable variation for a given category. This analysis of course assumes as a

convenience that the degree of overlap among vowel pairs is a bidirectional relationship, where the overlap is a product of the contribution of the variance from both categories. However, there is good reason to assume this may not be the case, and certain categories are more likely to vary and encroach on the boundaries of another vowel category than the other direction. For example, /u/ has greater variance than /i/ and is more likely to encroach upon the boundaries of /i/ as it shifts forward, but the reverse is not necessarily true. Pillai scores otherwise miss this directionality, by indicating overall degree of overlap between /i/ and /u/. The next section seeks to better understand the directions and variance of individual categories along specific cue dimensions in greater detail.

3 Part 2: Cue Specific Tendencies & Socio-Indexical Factors

Beyond the general principles of overlap, sociophonetic work offers numerous additional insights into how variation is socio-indexically structured. This section focuses specifically on insights regarding patterns of variation in F1 and F2, and the different ways socially structured variation shapes distributional properties. While there are many good reasons to treat vowels in multivariate space, with cues comprising a set, it is also analytically valid to examine cue specific tendencies within a set, and indeed this is where most sociophonetic analyses begin their descriptions. As will be illustrated in this section, the examination of F1 and F2 independently is motivated by evidence that vowel categories demonstrate unique distributional properties along specific dimensions (Van Hofwegen, 2013; Fruehwald, 2013, 2017) and follow regular patterns of variation along specific axes in accordance with social factors (Labov, 1994, 2001; Thomas, 2011).

Additionally, evidence from distributional learning literature demonstrates that people can attend to and learn patterns of individual cue dimensions even when they share covariance and utility as a set (Kruschke, 1996). These two facts combined suggest that examining F1 and F2 in multivariate space (as in previous analyses) may be obscuring structured variation along specific cues that language users learn. This section serves to uncover how variability along F1 and F2 may demonstrate socio-indexical structure across American English. To set the stage for the analyses, I will provide an overview of current work that motivates examining specific cue distributions. Much of this work is drawn from sociophonetic research on language change,

however the overall goal is not to address issues in language change itself but rather to highlight the specificity of socio-indexical structure and provide some initial expectations for cue specific variability among American English vowels. In addition, I will review work across other domains of linguistics that speaks more directly to the distributional properties of acoustic cues.

3.1 Background

Typological regularity of sound change patterns in American English has been an emphasis of work in sociophonetics and motivates the need to examine socio-indexical structure along specific cues for further refinement of hypotheses in perceptual learning. The typological patterns of sound change are drawn from dynamic synchronic and historical processes. While this dissertation does not seek to speak to the mechanisms or processes of sound change, the work discussed here illustrates patterns that may be replicated synchronically as cross-talker variability in American English, or in the phonetic fluctuations that may give rise to the patterns over time.

Labov (1994, 2001) has argued for several Principles of Vowel Shifts that constrain the patterns of variation seen across time and space. Specifically, three principles offer insights into vocalic variability for this study: Principle I, III, and Principle IV. Principle I states that back vowels typically front, which has the implication that back vowels are more likely to vary along F2 than F1. However, Thomas (2011) notes that this mostly affects /u/ and other back vowels may raise to fill in the gap, thus we may see vowel specific tendencies rather than overarching subsystems. Principle III indicates long vowels front and raise, and low vowels fall and back. In such cases talkers may show more multivariate changes in tense vowels but follow a systematic, albeit more complex, trajectory of change. Nonetheless, we might see that talkers are more likely to vary along one dimension more than the other in these cases. Finally, Principle IV: peripheral and non-peripheral vowels may swap positions, as demonstrated in cases such as /eɪ/-/ɛ/ reversal in the SVS. Principle IV is constrained by the 'Lower exit principle' which notes that as a vowel falls, it will eventually hit bottom and enter peripheral space when it reaches an /a/ value, further evidence of /a/ providing some degree of anchoring and stability in the vowel system. We might hypothesize that talker variability is likely to occur along specific cue distributions, with

variation along a single cue dimension being more (or less) likely for a particular category or sub-system.

Recent work has also suggested that variability along F1 and F2 may demonstrate distinct distributional properties for categories undergoing change or used for stylistic variation (Fruehwald, 2017; Van Hofwegen, 2013, 2017). Intra-talker variability may be much wider for certain cue dimension in categories undergoing change (e.g., /æ/ retraction in the CVS) than those that are relatively stable (e.g., /u/ fronting in the CVS; Van Hofwegen, 2013, 2017). Additionally, Van Hofwegen (2013) shows that variability in vowel categories may be non-normally distributed for those undergoing change arguing the edges of the multimodal distributions are the loci for sociolinguistic style. Contrastingly, vowel categories that are not involved in sound change appear normally distributed and are not subject to the same stylistic use or range of inter-talker variability. Similarly, Fruehwald (2017) demonstrates that despite having similar intra-talker ranges, there is a narrower range of inter-talker variation for men than for women in Philadelphia and the shifts in range are not necessarily tied to shifts in the average across groups. Thus, beyond just the overall degree to which variation occurs in a single cue dimension (e.g., F2 for both /u/ and /æ/), there may also be differences both in terms of the shape and parameters of individual cue dimensions that may demonstrate socio-indexical structure. In addition, we might expect that within-talker variation and between-talker variation may have different impacts on distributional properties, but generally these two analytical lenses have not been examined in parallel.

In addition to these broader phonological patterns, examining the shape and make-up of distributions is important for understanding learning and speech perception and provide insight into community patterns. The make-up of cue distributions and their influence on learning has been discussed more extensively in literature on distributional learning. While there is good evidence that listeners attend to sets of cues (e.g., F1 + F2 as we have discussed so far), there is also evidence that individuals can attend to specific dimensions among sets of cues (Kruschke, 1996). Additionally, as evidenced by the category variability effect (see e.g., Cohen et al., 2001), distributional properties among cues may influence listeners' categorization behavior. Individuals categorize ambiguous stimuli according to the distributional properties of the contrasts, such that ambiguous stimuli are more likely to be associated with the higher variance

contrast than the narrower category, despite the stimulus being closer to the central tendency of the narrow distribution (Cohen et al., 2001). Kapatsinski (2018: 133) suggests listeners may infer that sampling from a higher variability category is more likely to produce an outlier value, therefore listeners are more likely to extend the category past the experienced exemplars. This warrants examining specific cue distributions to better understand how talker variability is structured along specific axes, and how this further constrains perceptual learning.

In accordance with Bayesian models of speech processing, listeners may build generative models about within and across talker variation in response to the shape of distributions. Distributional learning can improve category discrimination when items belong to different statistical modes of the same category—suggesting that listeners attend to the distributional properties of a single dimension even when they are not necessarily indicative of separate categories. In particular, a unimodal production from the talker in exposure would lead listeners to build a generative model that the talker has one production target along that dimension. Listeners are sensitive to the fact that not all productions are going to exactly replicate the target, but this is largely inferred as a degree of random noise that is not relevant to meaning (see also Kapatsinski, 2018). In contrast exposure to a talker who has a bimodal distribution may appear to have two production targets, which is enough for listeners to generate a model of the cause of these two distinct targets. In such cases, tracking the multi-modality and conditioning factors allows listeners to learn, for example, a phonological pattern needed for producing allophonic variation, or other contextual causes.

Such a mechanism may be useful in the contexts outlined by Van Hofwegen (2017), where multi-modality is associated with different styles and sound change. When encountering speech sounds such as /æ/, individuals might build generative models about talkers with more multimodal productions to have multiple meaningful production targets, which provide socio-indexical information, and information about variants used in specific situations. Such a generative model would aid in an individuals' ability to replicate variation for their social goals, as well as aid in speech processing in future contexts. Further, the shape of the distribution may influence how listeners infer ambiguous stimuli that may fall along the edges of multi-modal cue distributions. In cases where the cue distribution is characterized by wider variability, listeners may have less confidence in the cue and down weight the importance of the cue for

categorization (Kleinschmidt & Jaeger, 2015). However, if the wide range of variability is structured by social factors, such as situation or group identity, listeners may orient toward the more variable cue to replicate the pattern for social goals. In addition, categories that have greater or a more multi-modal distribution in a single cue dimension and narrow or normal distribution in the other may demonstrate differential adaptation along specific dimensions as a result.

In light of these discussions, we can predict that variability is conditioned along specific cue dimensions for individual contrasts in production. For example, overall patterns of /u/ may demonstrate greater propensity for variability along the F2 axis as expected from the typological tendency for vowels to front (and not lower/raise). Of course, this may be demonstrated through between-talker variation where talkers' or dialects' central tendency may be distinguished along this axis. Alternatively, there may be greater within-talker variability, where there may be greater token variability along F2 as talkers adopt ongoing sound changes or contextually governed variation. Similarly, for categories like /æ/, within-talker variation may be greater and demonstrate multimodal distributions within-talkers as it is used for more stylistic variation. Finally, there is of course the possibility that the flat raw cue distribution captures token level variability that results from shared phonological conditioning (as discussed in Chapter 2). For example, /u/ variability along F2 may be likely to occur both within and across talkers as a function of pre-coronal conditioning of /u/ to front. These variable realizations may result in intra-talker variability (i.e., talkers' distributions) largely aligning with the tendencies of the overall dataset or their dialect areas and show less reduction and increased normality. In the following sections I will examine each of these levels in more detail, focusing on a subset of vowels.

3.2 Methods

3.2.1 Quantifying Distributions

In this section to better understand overall distributional properties of F1 and F2, I rely on descriptive statistics about the distributions. I examine the central tendency (here, the mean), standard deviation, skew (γ_1) , and kurtosis (κ) of F1 and F2 separately. Each of these measures provide details about dispersion of the data, which has strong implications for how listeners may

deal with variability along each cue dimension. Kurtosis describes how broad the distribution is relative to its center (i.e., mean) and the propensity of the distribution to have outliers and greater density in the tails. Lower kurtosis ($\kappa < 3$) can be an indication of bimodality with higher proportions of data within the tails of the distribution or greater propensity to outliers; higher kurtosis ($\kappa > 3$) is an indication of a more data concentrated near the center and lighter tails. Skewness captures the asymmetry in the distribution, with negative skewness indicating longer left tail (e.g., more extreme negative values) and positive skew indication longer right tail (e.g., more extreme positive values). Rules of thumb for kurtosis vary but following George and Mallery (2010) values of +/-2 of the normal distribution ($\kappa = 3$) are considered within the range of a normal univariate distribution. Rules of thumb for skewness follow that values between -0.5 and 0.5 are fairly symmetrical, and between 0.5 and 1 and -0.5 and -1 are considered moderately skewed and values greater than 1 or less than -1 are considered highly skewed. Normally distributed data has a kurtosis of 3 and a skew of 0.

While there are statistical models that may detect goodness of fit by comparing models with different types of unimodal distributions, these are likely to lack power on an individual talker basis due to lower sample size and may not capture distributional tendencies where bimodality may result in close modes and broad variance. Additionally, other distributional tendencies may be apparent outside of unimodality and bimodality. The more homogenous the distribution is the more likely it is to demonstrate low standard deviations and high kurtosis. On the other hand, the more heterogenous the distribution is the more it is likely to have higher standard deviation and lower kurtosis. As such, I use a combination of these descriptive statistics and visual inspection of distributions to evaluate cue distributions below. All measures were calculated in R (R Core Team, 2018), with skewness and kurtosis[5] measured using the moments package (Komsta & Novomestky, 2015) and other measures (i.e., mean and standard deviation) calculated using the base functions in R.

[5] Kurtosis is calculated using Pearson's measure of Kurtosis and is not reflective of excess kurtosis. Skewness is based on Fisher/Pearson moment coefficient of skewness.

3.2.2 Defining Groups

Following previous analyses, I measure each of these descriptive statistics over different socio-indexical groups conditioned on vowel category. Descriptive statistics were calculated over the entire dataset, again all tokens across talkers and dialect areas. Additional factors follow previous analyses, where the Dialect factor represents statistics calculated over all tokens and talkers of individual dialect areas, averaged over dialect levels. Dialect-specific patterns represent the individual dialect levels, as calculated over all tokens and talkers. The Talker factor represents descriptive statistics calculated over all tokens for an individual talker, and then averaged. Finally, Dialect+ Talker represents distributional properties over individuals' central tendency (i.e., mean) for a particular cue distribution. As such, the unit of analysis shifts from the distributional properties of a talker to their average behavior. Rather than considering token level variability which may vary within and across talkers, it provides a summary measure of their mean behavior on F1 or F2. Therefore, each descriptive statistic here is describing the distribution of talker means (e.g., skew of talker means of F1 for /a/). As such, this also captures some degree of hierarchical structure where parameterization of talkers occurs within the larger group structures and provides an estimate after accounting for intra-talker variability.

3.3 Analyses

3.3.1 American English (All Data)

Following the logic of the previous analyses, I will focus on describing the distributional properties more broadly across the entire dataset, and then move towards descriptions of the average properties across different socio-indexical factors (Dialect, Talker, Dialect + Talker) to illustrate the ways in which socio-indexical structure shapes distributional properties. I will primarily focus the descriptions on categories that are implicated in having non-normally distributed properties as a result of typological patterns (/u/ and /a/), sound change and style (here focusing on /æ/ and /ɔ/) and implicated in previous analyses in this dissertation (/eɪ/ and /ʊ/). I will then discuss the implications for inferential speech perception processes in the broader discussion (Section 3.4).

Figure 5.8 illustrates the probability distributions of Lobanov normalized F1 and F2 across vowel categories (all tokens, all talkers) with a point for the mean and credible intervals. Table 5.4 describes these distributions in terms of traditional descriptive statistics giving the mean, standard deviation, skew, and kurtosis. As depicted in Figure 5.8, across vowel categories there is greater variability in F2 than in F1, which is supported in the descriptive statistics in Table 5.4. Overall, most vowel categories fit within the normal distribution range for skewness (-0.5 – 0.5) and kurtosis (3 +/- 2). The exception to this pattern is /i/ F2 (κ = 5.14) and /ʌ/ F1 (κ = 5.02), which generally show lighter tails and greater density of data near the center. However, visual depictions of the density distributions in Figure 5.8 shows some indication that for certain vowel categories the distributions appear to be near bimodal and show greater variability. In particular, the back vowels /u/ (F2: σ = 0.87, κ = 3.13, $\gamma1$ = - 0.68) and /ʊ/ (F2: σ = 0.77, κ = 2.94, $\gamma1$ = -0.20) show greater variability in F2 compared to F1, with a moderate negative skew for /u/, suggesting a propensity for more extreme backed tokens. This aligns generally with the pattern of the tendency for back vowels to front and shows that there is greater variability across talkers and/or tokens with more conservative backed variants.

Additionally, /ɔ/ (F1: σ = 0.90, κ = 2.25, $\gamma1$ = -0.25) and /a/ (F1: σ = 0.64, κ = 3.92, $\gamma1$ = -0.30) tend to show greater variability across talkers and tokens with wider or near bimodal distributions in F1 compared to F2. While still in the range for normality, the lower kurtosis values of /ɔ/ align with the visual inspection of the probability distribution for F1. This broadly aligns with the tendency for talkers to be variable in terms of their low back vowel merger. Additionally, /æ/ shows greater variability along F1 (σ = 0.80, κ = 3.22, $\gamma1$ = -0.24) than F2 (σ =

0.58, κ = 0.36, $\gamma 1$ = 0.14), which generally aligns with the fact that talkers may raise /æ/ across some dialects (e.g., NCS, SVS).

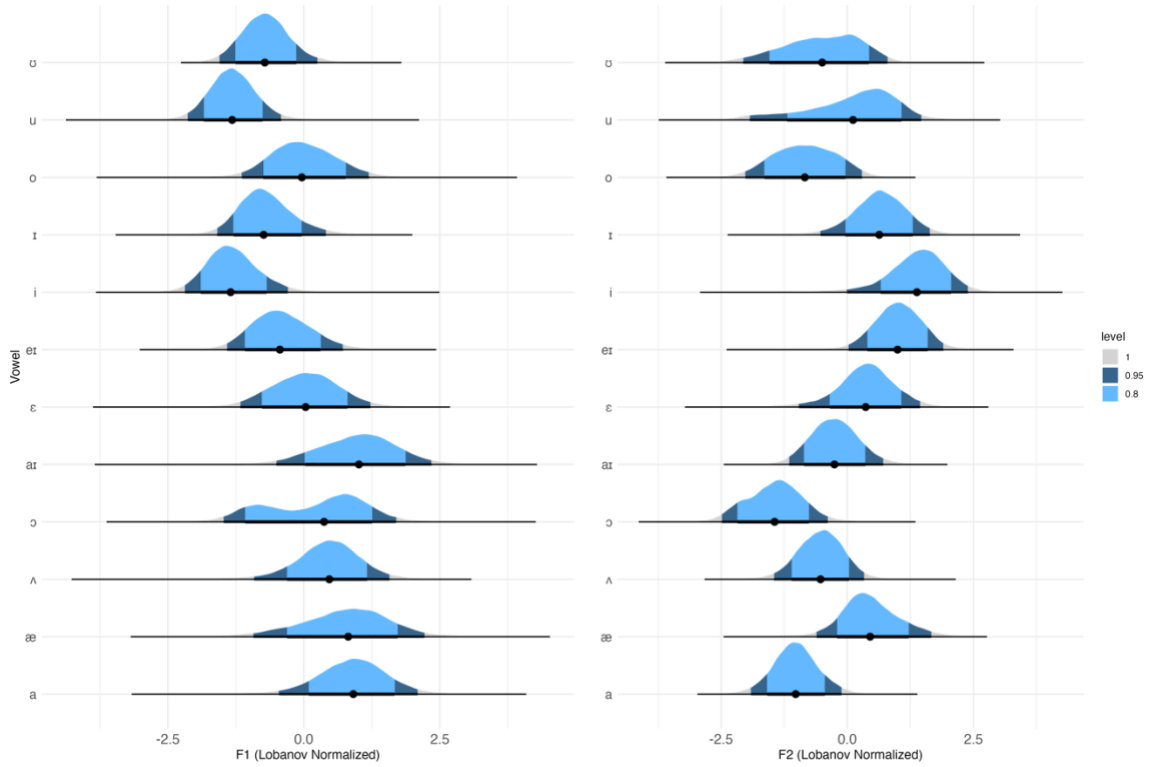

Figure 5.8 Probability distributions across Lobanov normalized F1 and F2 for individual vowel categories. Distributions are across all tokens and talkers in the dataset, fill values represent high density confidence intervals.

Table 5.4 Descriptive statistics for Lobanov normalized F1 and F2 across vowel categories, for the overall dataset distribution.

	Mean		**Standard Deviation**		**Skew (γ_1)**		**Kurtosis (κ)**	
Vowel	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**
a	0.89	-1.02	0.64	0.45	-0.30	0.07	3.92	3.33
æ	0.76	0.45	0.80	0.57	-0.24	0.14	3.22	3.57
ʌ	0.44	-0.53	0.62	0.46	-0.59	-0.15	5.02	3.42
ɔ	0.19	-1.44	0.90	0.55	-0.25	0.00	2.25	2.97
aɪ	0.98	-0.25	0.73	0.48	-0.19	0.10	3.39	3.12
ɛ	0.03	0.36	0.62	0.58	-0.03	-0.45	3.10	4.10

Table 5.4, Continued

	Mean		Standard Deviation		Skew (γ_1)		Kurtosis (κ)	
Vowel	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**
eɪ	-0.41	1.00	0.55	0.49	0.22	-0.45	3.20	4.84
i	-1.32	1.38	0.48	0.59	0.34	-0.8	3.61	5.14
ɪ	-0.70	0.63	0.5	0.55	0.38	-0.48	3.67	4.34
o	-0.01	-0.84	0.6	0.62	0.09	-0.05	3.27	2.59
u	-1.31	0.11	0.44	0.87	0.26	-0.68	4.19	3.13
ʊ	-0.70	-0.50	0.45	0.77	0.30	-0.20	3.75	2.94

These distributional patterns may arise from two possible sources. First, that talkers and/or dialects vary in terms of their average positions for these vowels, such that some talkers across groups are uniformly more fronted/backed or raised/lowered. Such a case may explain the bimodality of /ɔ/ (F1), which varies across varieties of English as a result of merger of the low back vowels. The second possibility is that bimodality or wider variability is indicative of talkers themselves being bimodal or widely variable indicating greater within-talker variation. The within-talker variation may be shared by talkers across American English, in the case of phonological conditioning, or may be idiosyncratic or stylistic. The latter might explain /u/, for example, as many talkers front in post-coronal environments and retain backed /u/ tokens in other phonological contexts (e.g., Labov et al., 2006). To better understand these different potential axes of between and within-talker variation, I will examine both the distributional properties of Talkers (Section 3.3.4) and the distributional properties of talkers' central tendency (Section 3.3.5), to better understand if the average position across talkers is driving the bimodality over and above token level variability. Put differently, do each of the modes represent different groups of talkers, or do individuals broadly demonstrate a similar bimodal (or wide) distribution of tokens? In the proceeding sections, I will compare the overall distribution parameters to the socio-indexical factors of Talker, Dialect, and Dialect + Talker to better

understand how distributional properties are conditioned on these factors. I will examine each category in terms of Lobanov normalized F1 and F2.

3.3.2 Dialect-Agnostic

Having established the baseline distributional properties, we now consider how dialects, on average (Dialect) distributional tendencies compare to the overall distributions. Figure 5.9 shows the descriptive statistics values for F1 across the entire dataset (blue circles, values from Table 5.4 above) and the average descriptive statistics for the Dialect factor (orange squares); Figure 5.10 shows the descriptive statistics for F2 across the same factors. These figures validate the overall patterns in Section 3.3.1 above that vowel categories are largely normally distributed, with skew and kurtosis values falling within a normal distribution range. Additionally, the Dialect factor approaches the more ideal skew and kurtosis of a normal distribution for both F1 and F2 across vowel categories.

Turning to the vowel categories of interest, we might expect that the broad range of variability in F2 for /u/ and /ʊ/ may be reduced when looking at tokens and talkers within a dialect area ('Dialect') if the distributional patterns are the result of cross-talker variability as a result of dialect differences. Such a change, however, is minimal as there is only slightly lower standard deviations in Figure 5.10 when we look at 'Dialect' averages for /u/ ($\sigma = 0.85$) and /ʊ/ ($\sigma = 0.74$). Additionally, there is still evidence of some negative skew across average Dialect factor for /u/ ($\gamma 1 = -0.43$), validating that there are generally extreme backed talkers and/or tokens within dialects as well. In other words, the propensity for outliers in the overall distribution cannot be accounted for by conditioning on dialect alone. This suggests either continued between-talker variation within dialect areas or some degree of within-talker variation that is maintained across group distributions.

Next, there is an increase in kurtosis for /eɪ/ F2 ($\kappa = 5.47$), showing a greater concentration of values near the center in F2 within dialect distributions compared to the overall dataset. However, there is a moderately negative skew in F2, indicating some more extreme centralized talkers and/or tokens. For the other vowel categories of interest, /a/, /æ/, and /ɔ/, we generally see that the Dialect factor shows a tendency towards more normality and lower standard deviations, while maintaining greater variability along F1 than F2. Similarly, /ɔ/

continues to demonstrate more heavy tailed distributions even when conditioning on dialects (κ = 2.60) suggesting that the bimodality observed across the overall data distributions is not accounted for when conditioning on dialect alone. Of course, these observations could be driven by some dialect areas where there is greater talker/token variation within the region. In the next section I will look at the dialect-specific tendencies before moving on to the other social factors (Talker and Dialect+Talker tendencies).

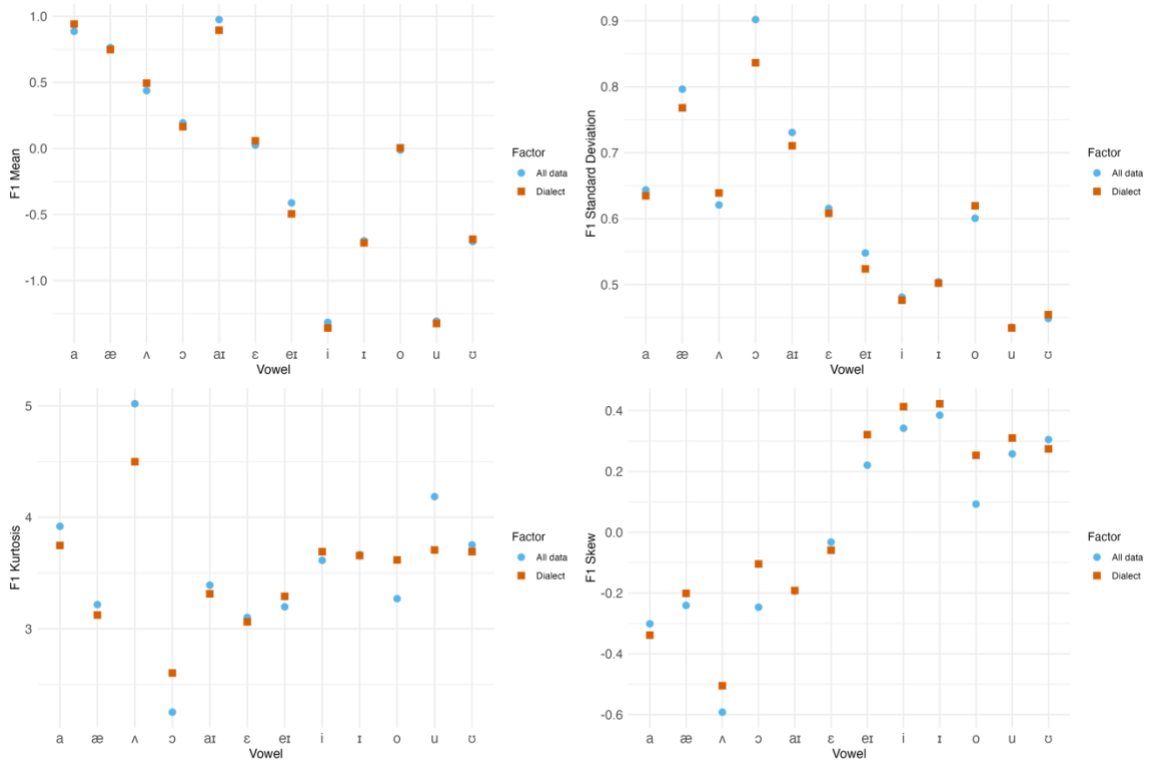

Figure 5.9 Comparison of descriptive statistics for Lobanov normalized F1 values for each vowel category distribution. Blue circles represent the overall distribution (all data, tokens, talkers). Orange squares represent the Dialect factor as the average of each descriptive statistic calculated over individual dialect levels.

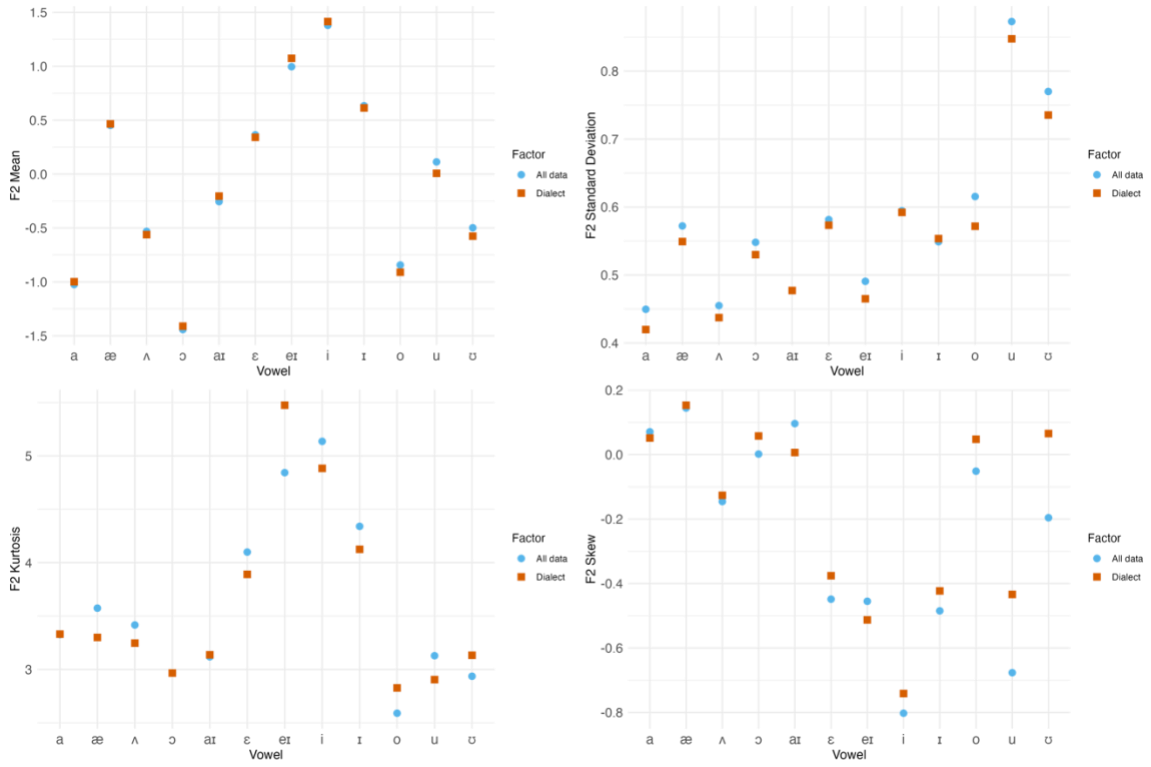

Figure 5.10 Comparison of descriptive statistics for Lobanov normalized F2 values for each vowel category distribution. Blue circles represent the overall distribution (all data, tokens, talkers). Orange squares represent the Dialect factor as the average of each descriptive statistic calculated over individual dialect levels.

3.3.3 Dialect-Specific Patterns

Turning to look at the dialect-specific tendencies, Figure 5.13 and Figure 5.14 show the probability density functions for each dialect area and vowel category to complement the descriptive statistics plotted in Figure 5.11 and Figure 5.12. Looking at the dialect-specific descriptive statistics in Figure 5.11 and Figure 5.12 we can see that generally all dialects and vowel categories approximate normal distributions according to accepted ranges of skew and kurtosis. Each dialect area demonstrates relatively similar trends for each of the critical vowels examined in this section, despite varying in central tendency for F1 and F2 across categories.

First, /u/ and /ʊ/ demonstrate more variability in F2 compared to F1 as demonstrated by higher standard deviation across dialects. Across dialect areas, the range of standard deviations

for F1 in /u/ (0.39-0.48) and /ʊ/ (0.41-0.59) are narrow, but the ranges in F2 are larger (0.73-0.96 for /u/ and 0.57-0.78 for /ʊ/) in addition to the overall tendency for F2 to be more variable. Kurtosis values indicate a normal distribution with lower values for some dialects indicating greater spread of probability into the tails and greater propensity towards outliers in F2. Visual inspections of the dialect distributions similarly show some indication of bimodality in /u/ and /ʊ/ for each dialect category, demonstrating token and/or talker variability within dialect areas for F2. In contrast, there is minimal variability in F1 for these same vowel categories, with dialect areas being relatively similar to one another with lower standard deviations and slightly higher kurtosis.

For /æ/ and /a/ we can see that there is generally a greater range of standard deviations across dialect areas and density near the center of the distribution for F1. In comparison, F2 shows less indication of variability for /æ/ and /a/ with lower standard deviation (σ = range /a/: 0.37-0.47; σ = range /æ/: 0.49-0.67) compared to F1 (σ = range /a/: 0.55-0.68; σ = range /æ/: 0.65-0.93) and normal kurtosis (approximately $\kappa = 3$) across dialects. This largely aligns with the trends observed thus far for both the overall data distributions (Section 3.3.1) and average Dialect factor (Section 3.3.4). Additionally, /ɔ/ shows greater standard deviation in F1 and lower kurtosis compared to F2 across dialects. In fact, the kurtosis for /ɔ/ is among the lowest (i.e., most outliers in the tails) of all the categories and is greatest for the South, demonstrating more heavy tailed distributions. Figure 5.13 visually supports this observation, which shows near bimodal distributions across several of the dialect areas, with the Midatlantic distinguished by a normal univariate distribution.

Finally, /eɪ/ shows normality across dialects in F1. However, F2 shows higher kurtosis with all dialects greater than 3, showing largely more mass in the center of the distribution and lighter tails. Additionally, the Midatlantic appears to be the most extreme and shows high kurtosis ($\kappa = 11.1$) and negative skew ($\gamma_1 = -1.76$) in F2, which may be the result of fewer talkers in the sample. These findings align with the overall dataset (Section 3.3.1) and the dialect-agnostic description (Section 3.3.2). Broadly speaking, cross talker variation is normally distributed for /eɪ/ but may show lower variability and more concentration near the central tendency within dialect areas along F2.

Overall, this shows that the high back vowels have greater variability in F2 than F1, aligning with Labov's (1994) principles, and overall, more likely as a function of constraints imposed by articulatory limits. On the other hand, the low back vowels tend to show greater variability in F1 than F2 across dialect areas, as might be expected from the merger of the two vowels, where both categories may vary along the same axis, and from articulatory constraints of low vowels. Visual inspection also demonstrates /ɔ/ F1 has some degree of bimodality within dialect areas, in contrast to /a/ which conforms to a univariate distribution, suggesting greater token and/or talker variability for /ɔ/ than /a/. The front vowels of interest are quite different from one another. In particular, /eɪ/ appears to be distinguished by central tendency across dialects and is coupled with somewhat more peaked distributions along F2 compared to F1, demonstrating greater regularity and concentration near the center for dialects. In other words, /eɪ/ is distinguished by both changes in central tendency and more regularity across tokens and talkers in F2, and normally distributed F1.

Additionally, /æ/ shows greater variability along F1 across dialects and lower variation in F2, which is somewhat in-line with the tendency for some regional varieties to raise /æ/ (e.g., NCS and SVS). However, given that dialect areas are also prone to retraction (e.g., LBMS) it's surprising to see lower variability along F2 across the overall dataset and for individual dialect areas. The greater variability in F1 may be driven, in part, by phonological environment, where /æ/ shows variable patterns of raising pre-nasally (Labov et al., 2006). As noted previously, of course the raw distributional patterns may largely be driven by token variability, which may be shared across talkers in a region in the case of phonological variation. Alternatively, it could be because talkers vary in their central tendency, creating two modes that encompass different groups of talkers. The next two sections will begin to disambiguate whether such variability is a function of between-talker or within-talker variation.

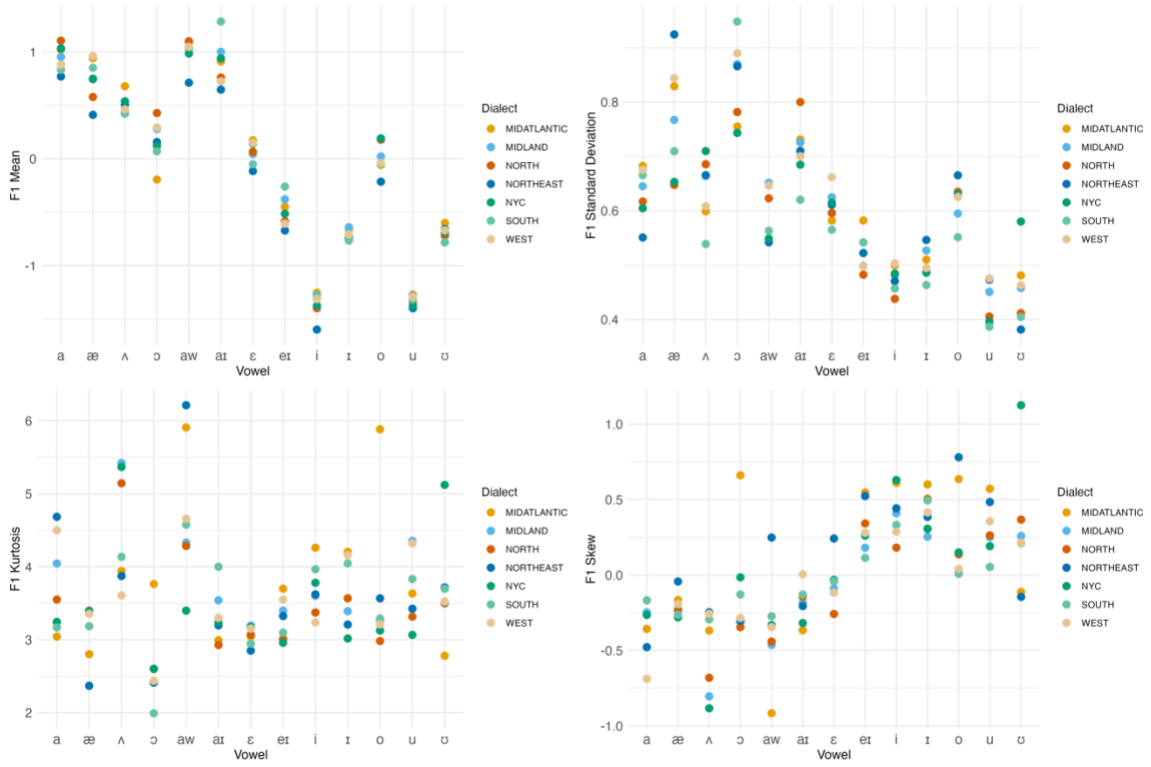

Figure 5.11 Comparison of descriptive statistics for Lobanov normalized F1 values for each dialect level, calculated over all tokens and talkers within each dialect area (e.g., all talkers and tokens from the South) conditioned on vowel category.

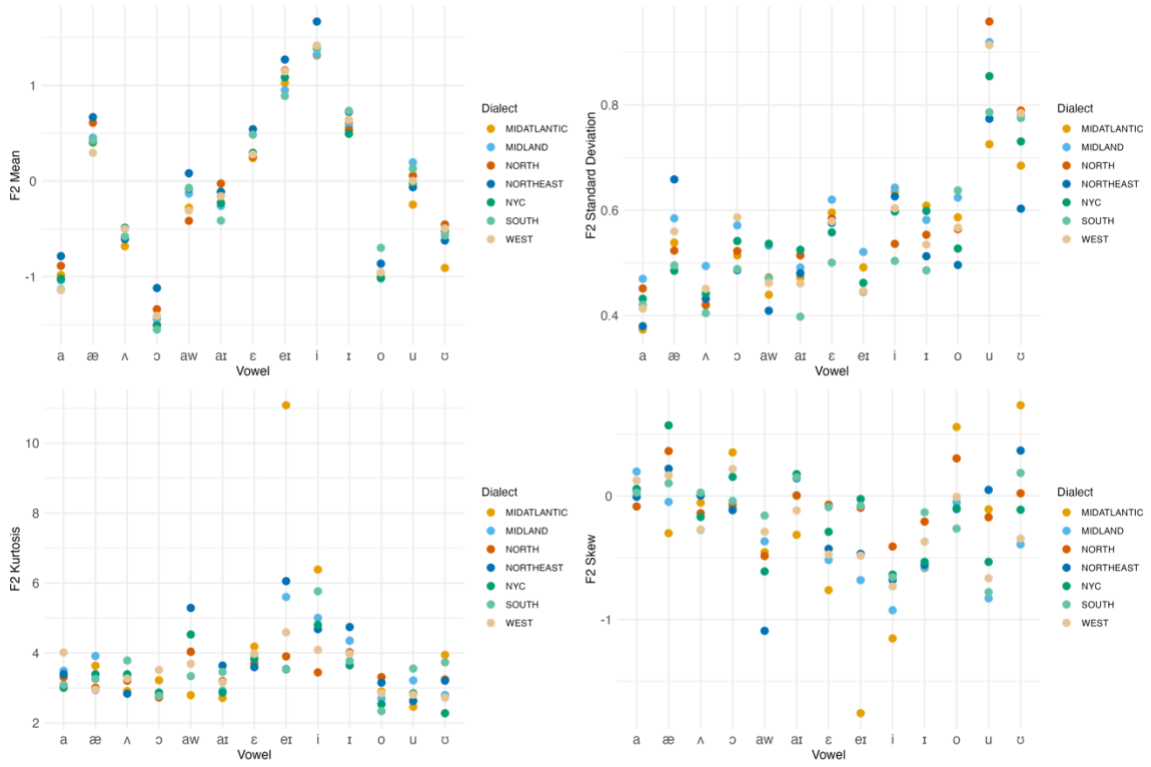

Figure 5.12 Comparison of descriptive statistics for Lobanov normalized F2 values for each dialect level, calculated over all tokens and talkers within each dialect area (e.g., all talkers and tokens from the South) conditioned on vowel category.

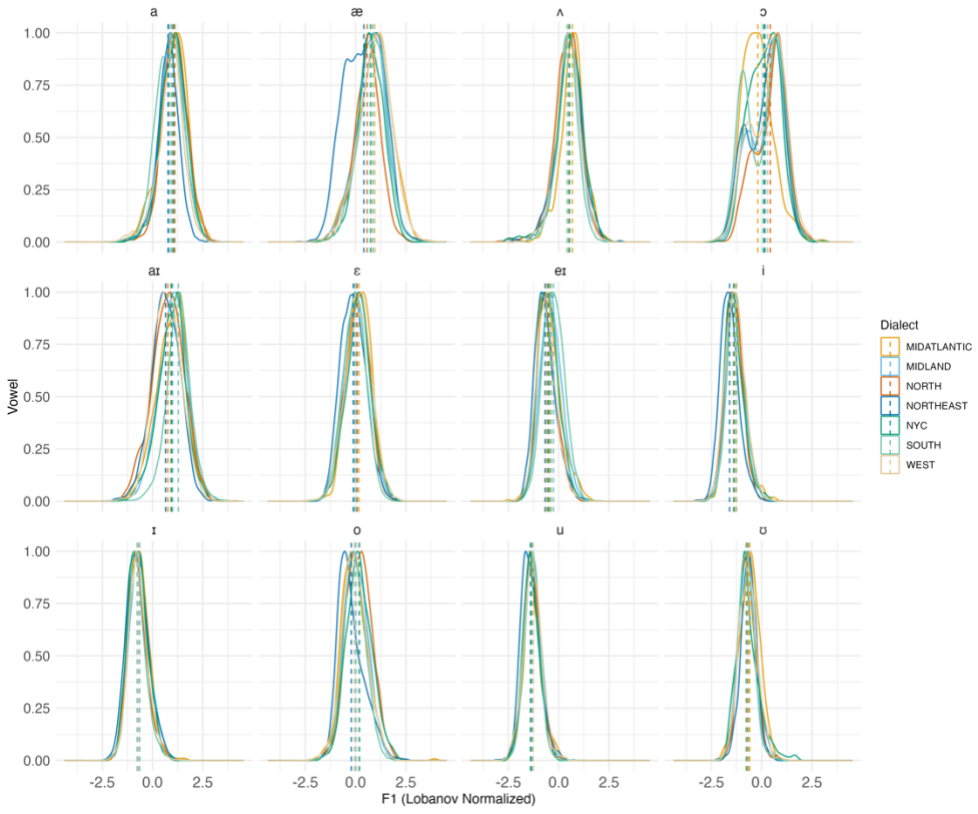

Figure 5.13 Probability density plots for Lobanov normalized F1 for each vowel conditioned on dialect area (indicated by color). Dashed lines indicate dialect area mean F1.

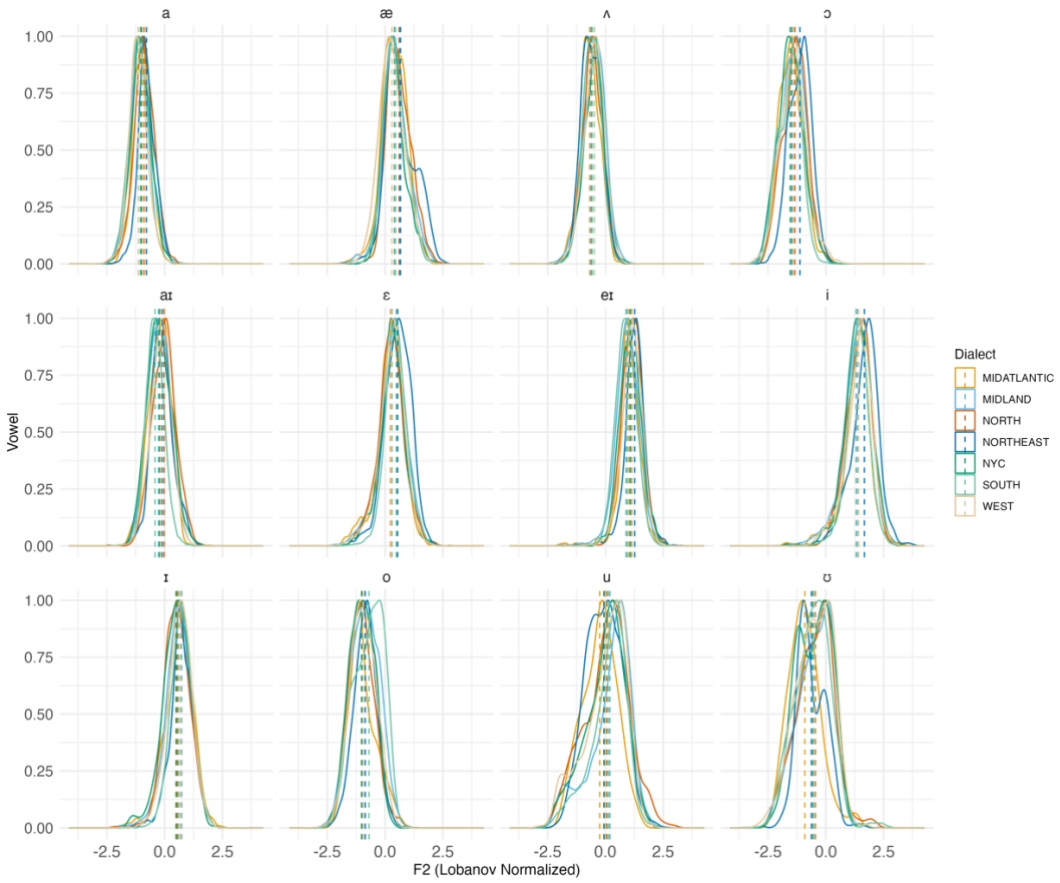

Figure 5.14 Probability density plots for Lobanov normalized F2 for each vowel conditioned on dialect area (indicated by color). Dashed lines indicate dialect area mean F2.

3.3.4 Talkers

Shifting from the dialect patterns, I now turn to look at how talker-specific distributions align with the overall dataset (Section 3.3.1) and the dialect-agnostic (Section 3.3.2) distributional properties. Figure 5.15 and Figure 5.16 provide the descriptive statistics of the overall dataset, the Dialect factor, and the Talker factor. The Talker factor represents an average of each descriptive statistic for each vowel category calculated for each individual talker's distribution. Across vowel categories we see that talkers' cue distributions are on average more normally distributed with lower standard deviations, less skew, and normal kurtosis, compared to the overall distributions and Dialect distributions. Such a finding confirms the expectations that talker-specific tendencies of cue dimensions are highly regular and normally distributed and

within-talker variability is less than cross-talker variability across this dataset. There are a couple of exceptions, which I will cover below.

Turning to the vowels of interest, we again observe that Talkers are on average more normally distributed for /æ/, /a/, /eɪ/, and /u/ compared to the overall and Dialect average. While /ɔ/ and /ʊ/ both show a somewhat lower kurtosis (κ = 2.51 and 2.43 respectively) and negative skew in /ʊ/ (γ_1 = -0.20), though still falling within the normal range. In addition, /ɔ/ F1 shows that talkers have lower kurtosis (κ = 2.3) and negative skew (γ_1 = -0.30) compared to the Dialect average which aligns more closely to the overall data distribution. These findings suggest that within-talker variability is similar to the distributional patterns of the overall dataset, where more data fall in the tails for /ɔ/ F1 and /ʊ/ F2 with a propensity for more backed tokens for /ʊ/ and higher tokens for /ɔ/. That is, talkers on average show similar token variability as the overall data distribution encompassing both between and within-talker variation, suggesting that within-talker variability is on par with between-talker variability for /ɔ/ F1 and /ʊ/ F2. Patterns in /ɔ/ may be due to the low back vowel merger in various parts of the U.S. and ongoing sound change may result in bimodal distributions within-talkers (Fruehwald 2017) in addition to the between-talker differences. In the next section I will examine how talkers' mean behavior varies along each cue dimension for the vowels of interest.

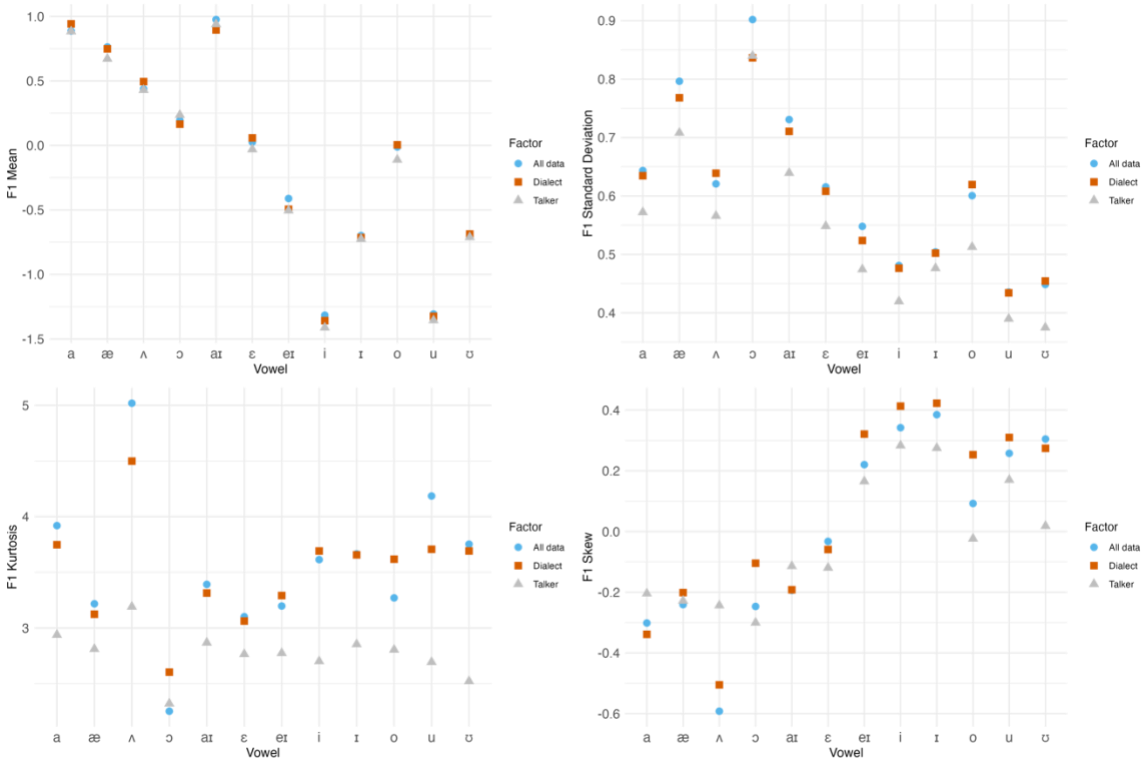

Figure 5.15 Comparison of descriptive statistics for Lobanov normalized F1 values for each vowel category distribution. Blue circles represent the overall distribution (all data, tokens, talkers). Orange squares represent the Dialect factor as the average of each descriptive statistic calculated over individual dialect levels. Grey triangles represent the Talker factor as the average of each descriptive statistic calculated over individual talkers (all tokens, raw distribution).

Figure 5.16 Comparison of descriptive statistics for Lobanov normalized F2 values for each vowel category distribution. Blue circles represent the overall distribution (all data, tokens, talkers). Orange squares represent the Dialect factor as the average of each descriptive statistic calculated over individual dialect levels. Grey triangles represent the Talker factor as the average of each descriptive statistic calculated over individual talkers (all tokens, raw distribution).

3.3.5 Talker Means

Figure 5.17 show the distribution of talkers' mean F1 and F2 for each vowel category across the entire dataset with the mean and high-density point intervals for each dialect area overlaid (made in R using the ggdist package, Kay, 2020). Figure 5.18 and Figure 5.19show the descriptive statistics for each vowel category and cue dimension's distribution of talker means. The green plus sign reflects the descriptive statistics of the distribution of means in Figure 5.17 and the orange triangle reflective of the descriptive statistics of talker means grouped by their dialect area. A caveat should be noted for this section. That is, mathematically, we expect distributions of means to be normally distributed and as the sample size increases (i.e., number of unique talkers) the standard deviation of the distribution of sample means will decrease. In other words, the distribution of means across the entire dataset is expected to have lower standard

deviation compared to individual dialect areas' distributions of talker means (as illustrated in Figure 5.17 and supported by statistics in Figure 5.18 and Figure 5.19). Similarly, dialect areas with smaller sample sizes will have larger standard deviations of talker means, which is supported by Figure 5.17.

Despite this, Figure 5.17 depicts some interesting trends which are further supported in the descriptive statistics in Figure 5.18-Figure 5.19. Broadly speaking, the trends align with expectations demonstrating that the means of talkers' F1 and F2 are largely more normally distributed with lower standard deviations. However, for /æ/ talker means vary in F1 with a greater density of data distributed in the tails ($\kappa = 2.50$, $\sigma = 0.32$, $\gamma 1 = -0.07$). On the other hand, the variability in /æ/ F2 has greater density in the center of the distribution ($\kappa = 4.58$) and but moderate negative skew ($\gamma 1 = -0.50$), illustrating a tendency towards talkers who have a propensity towards retraction. That is, for /æ/ talkers on average vary more along F1 than F2, but F2 is more prone to more extreme retraction among talkers. Contrastingly, /eɪ/, /a/, /ɔ/, /u/, and /ʊ/ show normal distribution of talker means, as expected, but reaffirm that the more bimodal distributions depicted in the above sections may be evidence of greater token variability within and across talkers, rather than different modes resulting from individual talkers and dialects uniformly patterning together near the peaks.

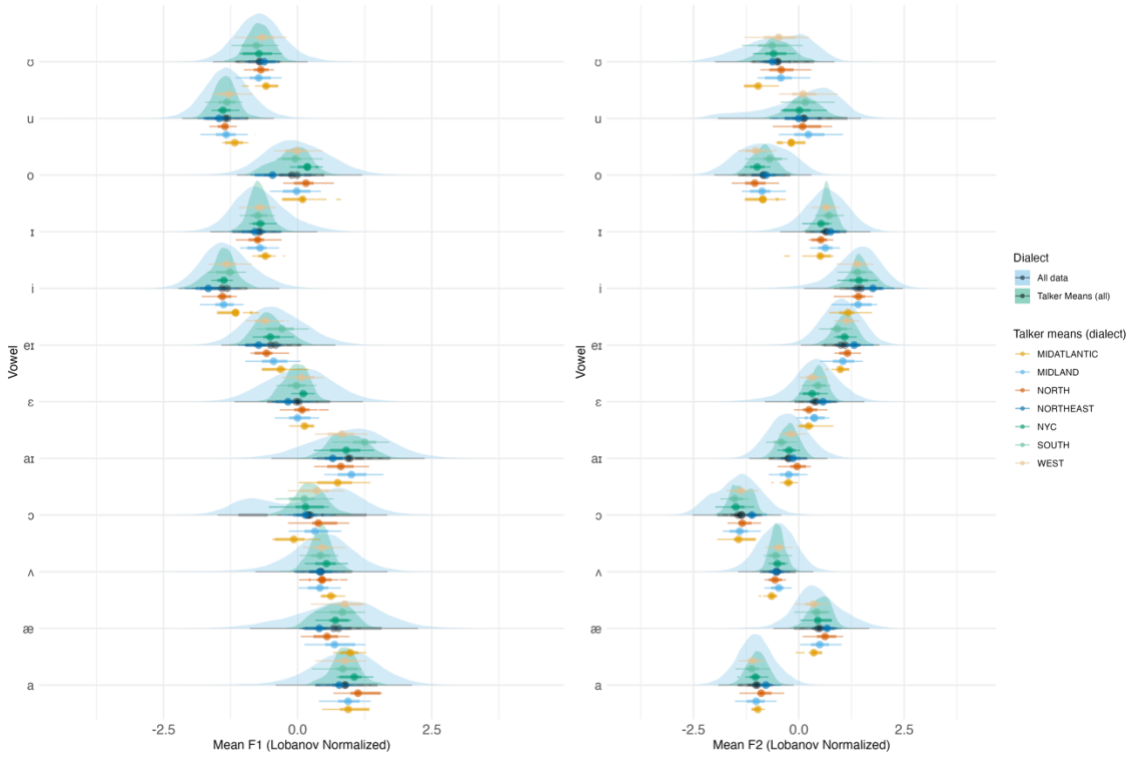

Figure 5.17 Light blue probability density curve shows the Lobanov normalized F1 and F2 distribution for each vowel category. The green probability density curve shows the distribution of talker means across the dataset for each vowel category. Point intervals show the mean (point) high-density interval (HDI). Thicker lines represent 66% HDI and thinner lines represent the 95% HDI.

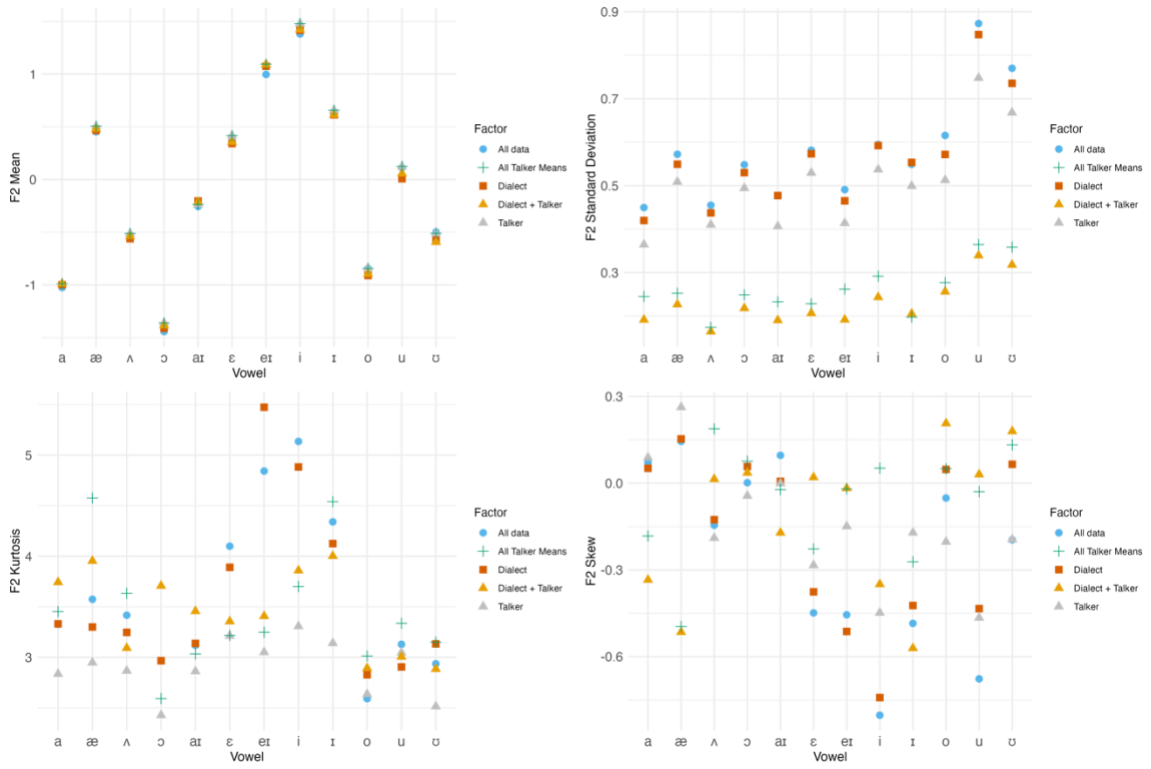

Figure 5.18 Comparison of descriptive statistics for Lobanov normalized F2 values for each vowel category distribution. Blue circles represent the overall distribution (all data, tokens, talkers). Orange squares represent the Dialect factor as the average of each descriptive statistic calculated over individual dialect levels. Grey triangles represent the Talker factor as the average of each descriptive statistic calculated over individual talkers (all tokens, raw distribution). Green plus values represent the descriptive statistic over the distributions of talker means. The orange triangles are grouped by dialect areas (e.g., the skew of talkers means in the South) and then averaged (e.g., the mean skew of talker means across dialects).

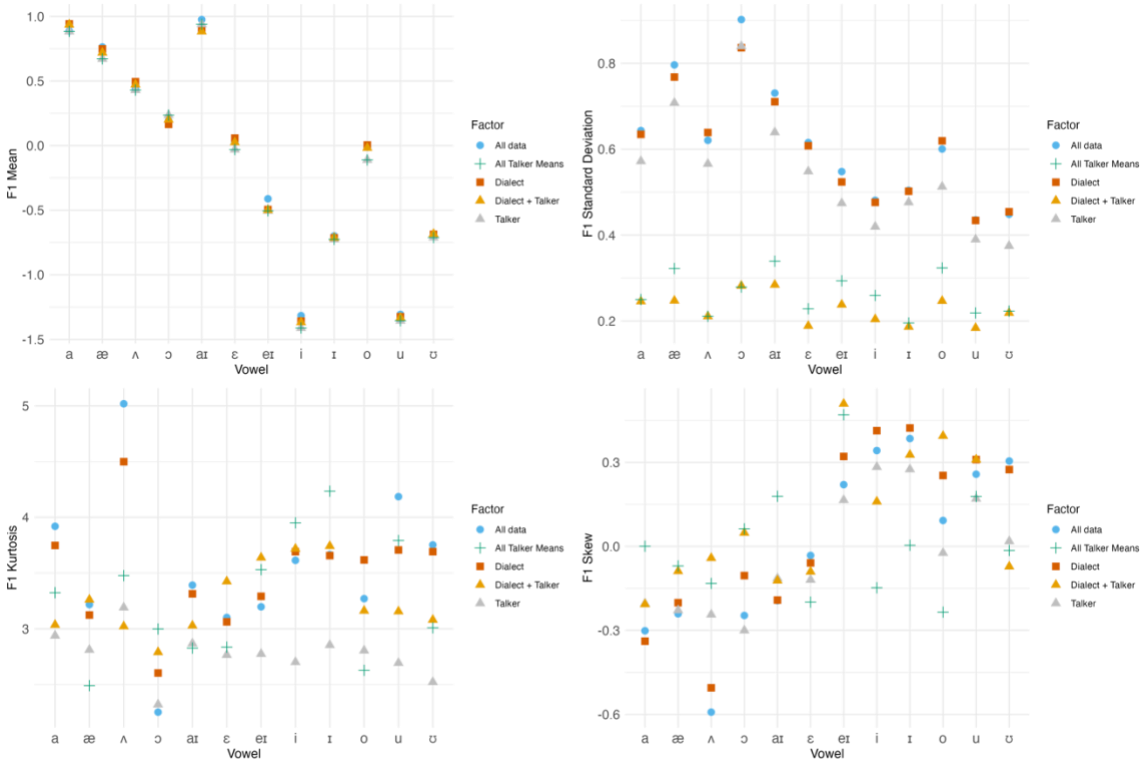

Figure 5.19 Comparison of descriptive statistics for Lobanov normalized F1 values for each vowel category distribution. Blue circles represent the overall distribution (all data, tokens, talkers). Orange squares represent the Dialect factor as the average of each descriptive statistic calculated over individual dialect levels. Grey triangles represent the Talker factor as the average of each descriptive statistic calculated over individual talkers (all tokens, raw distribution). Green plus values represent the descriptive statistic over the distributions of talker means. The orange triangles are grouped by dialect areas (e.g., the skew of talkers means in the South) and then averaged (e.g., the mean skew of talker means across dialects).

3.3.6 Interim Summary

Generally speaking, the results throughout this section validate expectations that talkers are regular in their cue distributions for both F1 and F2 across vowel categories, demonstrating lower standard deviations and normal skew and kurtosis on average. Examining the patterns of individual vowel categories, we can summarize several tendencies about between and within-talker variability. First, the analyses generally show that /u/ and /ʊ/ exhibit greater variability in F2 than in F1 with near bimodality occurring across American English. Dialect areas tend to adhere to the typological patterns observed in the overall data, with wider standard deviations in

F2 than in F1. However, in some cases there may be some dialect areas with skewed distributions in F1, suggesting some outliers in the observations. Additionally, talkers demonstrate wider intra-talker ranges in F2 for these categories, and greater skew towards backed tokens, suggesting other contextual factors may be driving the variability. Talker means overall appear normally distributed with similar ranges of variability in F1 and F2, suggesting the distributional patterns are not reflective of talkers who uniformly make up different modes (though are distinguished by central tendency). Overall, these patterns align with regional vowel shift expectations and typological regularities indicated by Labov (1994), whereby back vowels front, at least for the high back vowels.

On the other hand, the low back vowels demonstrate normally distributed cue dimensions for /a/ but greater variability in /ɔ/ F1 than F2, indicating greater cross-talker and within-talker variation in /ɔ/. Of note, we see that socio-indexical factors largely replicate the same patterns with greater variability in F1 than F2, and individual talkers mirror the patterns of the overall and dialect area distributions (i.e., near bimodality). The patterns in /ɔ/ may be partially resulting from the low back vowel merger changes in various parts of the U.S. In these cases, ongoing sound change may result in bimodal distributions within talkers (Fruehwald, 2017), as well as differences in central tendencies across regions. Similarly, examining /æ/, we saw that both talkers (Talker factor) and the overall data distribution demonstrate higher standard deviations in F1 (second only to /ɔ/) with low kurtosis, representing greater heterogeneity within and across talkers broadly. Finally, /eɪ/ shows greater regularity along F2 dimensions for the Dialect factor and the overall data distribution, with more density in the center of the distribution.

3.4 Interim Discussion

Overall, the patterns above illustrate the specificity of socio-indexical structure across specific cue dimensions that are otherwise missed when examining cross-talker variability in multivariate space (Chapter 4 and Section 2 above). First, categories that are not necessarily informative in multivariate space still demonstrate socio-indexical structure when examined along specific cue dimensions, and may reflect general typological tendencies (e.g., back vowels fronting). Second, categories that were informative of group-level information in Chapter 4 do not all show the same distributional properties of F1 and F2. That is, categories like /ɔ/ and /eɪ/,

which were informative in previous analyses, show different patterns such that variability is more like to occur in F1 for /ɔ/, leading to bimodality, but for /eɪ/ groups are distinguished by their means but have normally distributed data with similar properties across F1 and F2. Third, across all results dialect area distributions demonstrate greater homogeneity than both individual talkers' distributions and the overall dataset distributions. This pattern demonstrates that talkers' distributional properties align with the aggregate patterns of the group. That is, for certain vowel categories talkers demonstrate similar distributional properties even when the dialect is more non-normally distributed (e.g., /ɔ/). That is, different modes in the dialect areas' distributions are not necessarily indicative of groups of talkers who uniformly pattern together or of individual talker outliers, but of talkers themselves.

This prompts several hypotheses regarding how these patterns constrain perceptual learning and generalization. A hypothesis that comes from these data are that listeners may build expectations about a category in cue-specific ways. For categories like /u/ and /ʊ/, the greater variability along F2 within and across talkers may make listeners more flexible in adapting to novel variation in F2. The fact that talkers share this property across groups should allow listeners to infer group-driven variation and high probability of generalization to another talker. However, given the low range of variability in F1 for these categories, listeners may be more rigid in adapting to variation in F1, and if they learn the pattern they might do so in talker-specific ways by inferring this pattern was idiosyncratic. Listeners may also show different behavior in how specific or relaxed the pattern they learn is. F1 variability might lead listeners to learn a broader category relaxation and demonstrate greater flexibility to directions not experienced (e.g., generalizing from back vowel lowering to the same back vowel raising), while F2 variability may exhibit more targeted learning (e.g., this category only fronts). This is in-line with the observation of typological regularities in learning observed in Babel et al. (2021) where listeners who are exposed to typologically uncommon patterns of sibilant variation show general relaxation of categories compared to targeted learning in the direction of typologically common tendencies and may explain some of the different behaviors in learning novel chain shifts (e.g., Weatherholtz, 2015).

On the other hand, categories where variability in F1 and F2 are similar (e.g., /eɪ/ and /a/), listeners may demonstrate similar behavior regardless of the dimension of the novel shift.

However, this may be constrained further by cross-talker patterns and broadness of the variability. For categories that demonstrate relatively narrow variability, listeners may be more rigid in their expectations for category membership (e.g., /i/). Whereas categories with greater variability, listeners may be more flexible in adaptation, by accepting more outliers outside of their previous experience (e.g., /eɪ/). While this has not been specifically explored, this is counter to the arguments presented in Kataoka and Koo (2017) where they argue lower variability categories may demonstrate learning, and higher variability categories may demonstrate less learning. In their study they shifted a low variability category, /i/, towards a high variability category, /u/, and vice versa (i.e., /i/ backs, and /u/ fronts). Their results demonstrate asymmetrical learning where listeners learn novel patterns in the lower-variability target (/i/) but not in the greater variability target (/u/). However, the exposed shifts were conditioned on phonological context (pre-liquid), where coarticulatory pressures could plausibly result in backing of /i/ but would not result in fronting of /u/ (and indeed, is the opposite pattern that /u/ demonstrates in American English). Additionally, /u/ generally fronts in other contexts, which could also mean that if listeners did not have pre-existing knowledge of /u/ in pre-liquid contexts, they might expect that /u/ generally fronts across talkers and approaches /i/, which may put the exposure items generally within the boundaries of /u/ more broadly. This is all to say that it's not clear that range of variability alone is the driving force for retuning, a point I will return to in Chapter 6.

Finally, distributional shape poses an interesting set of questions for inferential processes, especially in cases where there is bimodality. As has been suggested by Kapatsinski (2018), it's possible that listeners build a generative model from bimodality that would indicate two production targets. For categories like /u/, /ʊ/, and /ə/ then, the bimodality (or near-bimodality) within and across talkers may result in listeners building a generative model around two production targets for the talker and the causes linked to these targets. This analysis doesn't uncover the causes of the bimodality of different categories, but different causal links to the bimodality may result in different listener behaviors. In some cases, these targets may be the result of phonological context (e.g., pre-coronal /u/ fronting) which may exist across all talkers in the same community, showing similar intra-talker variability. Listeners may extrapolate this pattern and recognize the conditioning contexts, which would allow them to replicate the pattern

in their own speech and adapt only when variation in those contexts is novel. Alternatively, if the pattern is the result of ongoing sound change, listeners may build a model around other socio-indexical factors that condition the change (e.g., gender or age), which will further guide predictions and inferences. Of course, an alternative is that listeners may not uncover the "true" distribution from which the data were generated and may infer normality with a wider standard deviation[6].

4 Conclusion

To conclude, this chapter has provided insights into the degree of specificity of internal category structure that occurs alongside between-talker variability. In Section 2 we saw that acoustic overlap across vowel pairs largely reflects separation along articulatory dimensions, such that talkers generally show less acoustic overlap across category pairs that differ in dimension (e.g., front/back, high/low). In addition, we see that the separation of /æ/ and /a/ remains stable across talkers and dialect areas, validating hypotheses about the special status of these two vowel categories more generally. Similarly, acoustic overlap is greatest among vowel pairs in the middle of the vowel space where misidentification is more frequent. Accounting for talker-specific and dialect area tendencies reduces the acoustic overlap for vowels near the center of the vowel space, suggesting that listeners may benefit from tracking cross talker variation for these categories, despite not appearing as informative in Chapter 4. Finally, the degree of acoustic overlap among categories may broadly provide listeners with expectations about the relative boundaries of variability for a given category. At the simplest form, this may provide listeners with the relative probability that a given item belongs to one category or another when the item is perceptually ambiguous. In other cases, listeners may be less likely to learn or generalize talker-specific tendencies where ambiguous tokens encroach on a category boundary that shows greater acoustic separation (e.g., /æ/-/a/). Future work should continue to examine how the relationship between contrasts constrains perceptual learning and generalization.

[6] /ɔ/ overall poses a secondary issue as it is merged with /a/. However, evidence shows listeners are sensitive to the distribution of /ɔ/ even when they are themselves merged talkers, as evidenced by cases of near merger and accommodation to unmerged talkers (see Chapter 2).

In Section 3, the distributional properties of specific cue dimensions highlighted the interplay of between and within-talker variation. Overall, this section showed that within category variability demonstrates typological tendencies which are indicative of both between-talker variation and within-talker variation, including the tendency for (high) back vowels to front, with variability largely occurring in F2. Similarly, the distributional tendencies of /ɔ/ and /æ/ demonstrate greater variability along F1, which is maintained across dialects and within talkers. On the other hand, /eɪ/ and /a/ demonstrated more normally distributed data regardless of socio-indexical factor with a tendency for greater regularity along F2 for dialect areas. These results validate, to some extent, the fact that talker variability is more likely to show greater variability when categories are associated with sound change or style (e.g., /ɔ/ and /æ/). Though, the distributional shapes and tendencies vary substantially between categories, which may suggest that different mechanisms may underly these tendencies, at least synchronically for these data. Regardless, the synchronic variability observed in this section highlights that listeners may form expectations about between-talker variation along specific cue dimensions (e.g., F2 for back vowels) and also may expect greater variation within a talker for a specific dimension (e.g., F1 for /ɔ/).

Overall, this chapter has highlighted that the more holistic treatment of variability of vowel categories may miss more fine-grained predictions about how listeners respond to novel talker variation. While this dissertation doesn't test the observations or predictions in this chapter, I will return to some of these points in explaining perceptual behavior of participants in the experiment of this dissertation (Chapter 6) and in broader discussions in Chapter 7.

CHAPTER 6:
SOCIO-INDEXICAL INFERENCE IN PERCEPTUAL LEARNING

1 Introduction

In the previous two chapters, socio-indexical structure across vowels in production was examined under various analytic scopes. Several questions were raised about the how these different analytic scopes correspond to different predictions about listeners' behavior in perceptual learning. Chapter 4 in particular highlighted one primary question regarding whether listeners exhibit different perceptual learning or generalization as a function of asymmetrical properties of socio-indexical conditioning in production. Driven by the analyses in Chapter 4, I investigate whether vowels that are distinguished by the types of socio-indexical conditioning demonstrate different listener behaviors in perceptual learning.

The focus of this chapter is individual talker identity and dialect background as the central social factors of interest. Grounded in listeners' a priori knowledge of the categories as inferred from the corpus analysis in Chapter 4, these different levels of social factors are hypothesized to prompt listeners to draw on different social causes underlying variation (as discussed in Chapter 4 and in detail below). In the context of perceptual learning, I hypothesize that vowel categories that show distributional patterns conditioned on dialect groups (i.e., heterogeneity between groups) and illustrate homogeneity among talkers within groups, will demonstrate more robust learning and cross-talker generalization. In this chapter I will examine /eɪ/ as a category representing this pattern and refer to it as *dialect-informative* throughout. On the other hand, categories whose distributions are not conditioned on dialect and show higher talker heterogeneity within dialect areas, but remain informative at an individual talker level, may demonstrate talker-specific learning. Here, I will examine /ʊ/ as the category representing these patterns and refer to it as *talker-informative*.

The experiment presented in this chapter uses a lexically guided perceptual learning paradigm to test how listeners' perceptual learning and generalization behavior may differ across these two vowel categories. After an initial pre-test categorization task, the experiment exposes listeners across two conditions to a novel vowel shift with an ambiguous phone between /eɪ/ and

/ʊ/ with the same talker from pre-test. In the /eɪ/-Biased condition, listeners hear ambiguous /eɪ/-/ʊ/ phones embedded in a lexically disambiguating context that biases them towards /eɪ/ interpretations (e.g., p[?]stry → p/eɪ/stry). In the /ʊ/-Biased condition, listeners hear the ambiguous phones embedded in a lexically disambiguating context that biases them towards /ʊ/ interpretations (e.g., h[?]king → h/ʊ/king).

Learning is evaluated based on whether listeners' categorization boundaries between /eɪ/ and /ʊ/ continua shift from pre-test to post-test for the same talker in the direction of exposure (i.e., learning). Generalization is evaluated based on listeners' categorization boundaries of the other talkers (either male or female) between the two conditions. Contrary to learning, generalization is evaluated more broadly as the extension of listeners' behavior from the exposure talker to another talker, regardless of whether the change in behavior is in the direction of the exposure.

In both the conditions I predict that from pre-test to post-test, listeners will learn the talkers' atypical productions thereby demonstrating a shift in the boundary between /eɪ/-/ʊ/ from pre-test to post-test. In the /ʊ/-Biased condition, an effect of learning is observed if at post-test, listeners provide more /ʊ/ responses than they did at pre-test, effectively shifting their categorization boundary towards the /eɪ/ end of the continuum. In the /eɪ/-Biased condition, an effect of learning is observed if at post-test listeners report more /eɪ/ responses than at pre-test, effectively shifting their category boundary towards the /eɪ/ end of the continuum. These two conditions, however, may demonstrate different degrees of learning, with more robust learning (i.e., greater magnitude) for the /ʊ/-Biased condition compared to the /eɪ/-Biased condition as a function of the category being more strongly conditioned on individual talkers (see Kleinschmidt, 2019). Finally, I predict that listeners will generalize the pattern in the /eɪ/-Biased condition to both the female talker and the male talker. However, in the /ʊ/-Biased condition, I predict that listeners will not generalize the pattern to either talker, or it will be limited to the same gender pair (i.e., the female talker). Thus, the two conditions are predicted to differ primarily by listeners' generalization behavior, where the dialectally informative category (/eɪ/) is likely to promote cross-talker generalization and the /ʊ/ condition is not. However, the socio-indexical asymmetry may also promote different magnitudes of learning of the exposure talker's pattern as well. Below I will outline the motivation and theoretical grounding for my hypotheses

in more detail focusing on ideal adapter models (Section 1.1), followed by a more in-depth description of the two vowel categories (Section 1.2). Then, I will give an overview of the experimental methods and results, and close with a discussion and conclusion.

1.1 Motivation

The hypotheses above are motivated by theoretical and computational work in ideal adapter models accounting for asymmetries across contrasts in perceptual learning and generalization (Clayards et al., 2008; Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008). Perceptual learning across contrasts has been demonstrated to be influenced by several factors including whether the exposure shift has an attributable cause (e.g., a pen in the mouth; Kraljic et al., 2008), stimulus attention (McAuliffe, 2015; McAuliffe & Babel, 2016), opposes the typical phonetic range of the contrast (Sumner, 2011), is phonologically driven (e.g., /s/ → /ʃ/ in /street/; Kraljic et al., 2008), or typologically uncommon (Babel et al., 2021), and the degree of variability in exposure (Sumner, 2011; Theodore & Monto, 2019; Wade et al., 2007). Correspondingly, cross-talker generalization has been shown to be influenced by the acoustic (Reinisch et al., 2014; Xie & Myers, 2017) or perceptual (Reinisch & Holt, 2014) similarity of the talkers, contrast type (Kraljic & Samuel, 2006), and variability in exposure (Babel et al., 2020; Sumner, 2011). Ideal adapter models attempt to account for the variable outcomes of learning and generalization through a process of inference under uncertainty, whereby listeners draw on previous knowledge to infer the underlying cause of variability and guide their perceptual learning behavior. This experiment attempts to account for asymmetries in cross-talker generalization under this theoretical model with a guiding hypothesis that different higher-order socio-indexical links to vowel category variability may result in asymmetrical perceptual learning and generalization.

In such accounts, listeners' previous experience aids in a priori inference about the underlying cause of the variability they experience which in turn motivates listeners' adaptation behavior (Clayards et al., 2008; Jongman & McMurray, 2017; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). During perceptual learning tasks, listeners integrate the current perceptual input with their prior experience at multiple timepoints. In such tasks the underlying cause is not typically directly observable, so listeners can only infer the cause from multiple

potential sources. Liu and Jaeger (2018) describe these contexts as *causally ambiguous*, due to the competition between potential causes and the absence of disambiguating contextual information. If listeners are provided disambiguating evidence for the cause of the pattern (e.g., a pencil in the speaker's mouth), the context is *causally unambiguous*.

During inference in causally ambiguous contexts, listeners allocate credibility across alternative explanations for the cause of the perceptual event. Initial allocation of credibility (i.e., before exposure to a novel shift) elicits listeners' a priori beliefs about which contrast is more likely to have caused the experienced input (i.e., the phone belongs to /eɪ/ or /ʊ/). When listeners are subsequently exposed to a novel shift, they must then allocate credibility to inferences about the underlying cause of the novel variation: Is it characteristic of the speaker, their social group, or some other incidental cause (e.g., a pen in the mouth)? The more likely a cause of the observed input will be present with future encounters from the same talker, the more fruitful it is to store and represent this pattern as a talker-specific pattern. At post-test, listeners then integrate their prior beliefs about the acoustic quality of the category with the knowledge of the acoustic quality of the categories from exposure, resulting in a posterior belief. Listeners then reallocate credibility during post-test categorization of the perceptual input based on whether the inferred causal model indicates the pattern is characteristic of the speaker. If the pattern is inferred to be caused by speaker specific characteristics, then listeners learn the pattern of exposure. On the other hand, if the pattern is inferred to be incidental and not characteristic of the speaker, listeners will not demonstrate learning.

Recent work incorporating socio-indexical structure into ideal adapter models suggest socio-indexical conditioning of a contrast influences listeners' inferences and behavior. In such cases, listeners may infer that socio-indexical factors are a potential cause for atypical pronunciation patterns (Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). If listeners have prior beliefs from experience that a contrast doesn't vary across individuals, then socio-indexical causes are unlikely to be inferred, and listeners are unlikely to interpret the underlying pattern as representative of the speaker's identity. As a result, they may remain rigid in their category representations, showing less remapping for the talker and no evidence of cross-talker generalization (Kleinschmidt, 2015). On the other hand, if a contrast is conditioned on grouping factors (e.g., dialect, gender, etc.) then there is a greater likelihood that

listeners infer the underlying cause of any novel variation to be conditioned by socio-indexical and representative of a speaker's identity. Because variability is conditioned not only on individual talkers, but also their groups, it's likely the variation could be leveraged for future speech processing tasks with other talkers, ultimately resulting in greater flexibility and generalization across talkers.

As discussed elsewhere (see Chapter 2), vowel categories that exemplify a circumstance of high socio-indexical causes of variability are hypothesized to have increased likelihood for learning and generalization. Discussions about the role of vowels and socio-indexical structure in ideal adapter models often assume a holistic perspective of vowels, whereby listeners draw on prior experience with the raw distribution of joint acoustic cue variability (i.e., F1 and F2) to infer the most likely interpretation (e.g., the speaker said *bat* and not *bet*) and the cause of any novel variation (e.g., social background). Yet, it's unclear whether listeners draw inferences about socio-indexical causes uniformly across the vowel space. One possible prediction is that each vowel category has equal potential for the same degree of socio-indexical inference. As such, listeners may have broad expectations that vowels are highly variable across talkers as a result of their dialect backgrounds. In turn, listeners' prior belief is that novel variation has a higher likelihood of being caused by dialect background and is therefore characteristic of the speaker and likely to be encountered again with other speakers. In turn, listeners learn novel variation regardless of vowel category, and generalize to similar talkers.

In terms of generalization, we may predict asymmetric generalization patterns across novel talkers based on the perceived talker characteristics and their social groups. Generalization is defined in this dissertation as an extension of the listeners' updated beliefs and shifts in categorization, regardless of whether it is in the direction of the exposure shift. That is, the change in listener behavior from pre-test to post-test of the exposure talker is extended to other novel talkers whose production patterns they have not experienced. When listeners infer that two talkers are from the same social group, listeners should be more inclined a priori to generalize from one talker to the other. However, if two talkers are inferred to be from groups that differ, listeners should be less likely to generalize. In the case of vocalic variation, listeners may infer that talkers are from different groups when they differ by gender, due to the overall gross differences in acoustics between the two groups (Kleinschmidt, 2019). For the experiment at

hand, that would indicate generalizing from the exposure talker to the second female talker, but not the male talker. On the other hand, there is a possibility that generalization of vowels may occur across gender pairs as listeners infer the talkers are the same broader dialect area and be inclined a priori to extend the pattern across gender pairs. In this case, the pattern is extended from the exposure female talker to both the novel talkers.

An alternative prediction, however, is that listeners make more fine-grained inferences depending on whether the individual vowel category tends to be more variable by social group (e.g., dialect) or individual talkers (i.e., idiosyncratic), rather than holistically at the 'contrast' level (i.e., all vowels behave in X ways). In such cases, listeners may demonstrate asymmetrical generalization, such that vowel categories that tend to vary more by dialect are more likely to show flexibility in adaption and generalization than categories whose variability is not typically caused by group membership. On the other hand, when a contrast is likely to vary across talkers but not conditioned on group membership, prior experience with past talkers is predicted to be less informative. As such, listeners are predicted to adapt more quickly and robustly (i.e., sharper shifts in boundaries) for contrasts that vary across talkers but do not show evidence of social conditioning. Subsequently, listeners are predicted to show talker-specific learning and no generalization to other talkers, or generalization is restricted to same gender talkers who are similar acoustically. This is the primary prediction tested in this experiment. As such, it is predicted that for /ʊ/ listeners will adapt quickly but restrict generalization. On the other hand, for /eɪ/ listeners are predicted to adapt and generalize to other talkers. However, given the difference in socio-indexical condition, the magnitude of the shift in learning may be greater for /ʊ/ than for /eɪ/. Listeners are expected to adapt more quickly and completely when talker identity is highly informative (*talker-informative*) because prior experience with other talkers will be less relevant a priori. On the other hand, listeners may initially approach the task with more a priori expectations based on prior experience with talkers in the case of /eɪ/ which may heighten uncertainty and slow adaptation.

There is limited work in perceptual learning of vowels, leaving many open questions about the accuracy and nuances of predictions made by ideal adapter models. Much of this work has not measured changes in phoneme categorization as is common in perceptual learning of consonants (e.g., Norris et al., 2003, Kraljic & Samuel, 2005, and others), and instead has

evaluated the adaptation to accents more broadly assessed through lexical decision endorsement rates after exposure to a story in a novel accent (Babel et al., 2019; Maye et al., 2008; Weatherholtz, 2015). Such work has illustrated that listeners learn cross-category remapping of vowels (e.g., /ɪ/ → /ɛ/) and may generalize to other phonologically related shifts (e.g., back vowels lowering) but generalization depends on the direction of the shift in exposure. For example, listeners appear to generalize from a novel front vowel lowered system (e.g., /ɪ/ → /ɛ/) to a phonologically similar shift of back vowel lowering (Maye et al., 2008; Weatherholtz, 2015). However, listeners don't generalize from a front vowel raising pattern to a back vowel raising system (Weatherholtz, 2015). These finding suggest some limitations to how listeners adapt to vowel shifts, such that some shift directions result in targeted learning (i.e., only learning the exposure shift) while others result in more global learning (i.e., generalization to phonologically related shifts). As discussed in Chapter 5, one reason for this may be that listeners have expectations from more common typological patterns thereby making some directions more marked and more difficult to learn (see also Babel et al., 2021). However, whether learning in individual vowel categories as opposed to larger 'accents' show constraints on learning depending on the directions of shifts is still an open question. While this experiment does not address this question directly, the results of the experiment may be elucidated by such an asymmetry.

Most relevant to the experiment in this chapter, other work examining perceptual learning through phoneme categorization tasks demonstrates listeners learn ambiguous patterns of individual vowel shifts (Chládková et al., 2017; Franken et al., 2017; Kataoka & Koo, 2017; McQueen & Mitterer, 2005). The broader accent learning described above alongside learning of ambiguous vowel shifts generally supports the claim within some ideal adapter models that vowels in general have a high degree of flexibility possibly driven by listeners' knowledge of socio-indexical structure. However, Kataoka and Koo (2017) demonstrate an asymmetry in learning across vowel categories in a group of American English listeners. In their work, after training, listeners learned a novel shift of /i/ → /u/ but did not learn a complementary shift of /u/ → /i/. The authors argue the asymmetry is driven by the properties of the individual vowel categories, such that the variability of each category is inversely correlated with category flexibility. Specifically, more variable categories (e.g., /u/) are less malleable and exhibit limited

perceptual learning while, on the other hand, lower variability categories (e.g., /i/) are more malleable. However, the authors do not differentiate between socially conditioned variation across talkers and other potential causes of variability. Furthermore, the shift of /u/ may have not been particularly novel to listeners since /u/ fronting is common in American English, which makes it difficult to evaluate whether "overall variability" of the category is truly the pattern driving the null effect of the /u/ shift verses whether the listeners already have experience with a shifted /u/. Consequently, it's unclear whether the results are evidence against the more holistic position identified in ideal adapter models whereby vowels should demonstrate greater learning, or whether listeners already had expectations for the pattern a priori and just received no *new* information that required them to update their beliefs.

Similarly, there is a paucity of literature examining cross-talker generalization of vowel shifts and this limited work has only examined large-scale shifts across the entire vowel space. Cross-talker generalization in perceptual learning in general is argued to rely on the acoustic (Reinisch et al., 2014; Xie et al., 2018; Xie & Myers, 2017) or perceptual (Reinisch & Holt, 2014) similarity of the talkers. However, Weatherholtz (2015) demonstrated that listeners generalize across talkers when exposed to novel cross-category remapping of vowel shifts, regardless of the acoustic similarity between the talkers. The acoustic similarity of the talkers was operationalized by the gender of the talker. Due to the gross differences of acoustic properties between male and female talkers, the same gender pair reflected acoustically similar talkers and the different gender pair reflected dissimilar talkers, an approach used in defining acoustic similarity (e.g., Kraljic & Samuel, 2006). The exposure talker was a woman, and the generalization talkers were either a same gender pair (i.e., woman) or a different gender (i.e., male). The generalization findings in Weatherholtz (2015) are in line with ideal adapter predictions that listeners generalize vowel patterns across talkers theoretically resulting from inferences about higher-order socio-indexical structure. Listeners indeed may have inferred the pattern (i.e., 'accent') was socially caused based on prior experience with complex vowel patterns across dialects, and thus more likely to extend the pattern to new talkers. While the overall indexical pattern is one potential explanation, a secondary explanation is that low level perceptual factors drive the generalization behavior. In particular, it's plausible the acoustic similarity of the talkers did not map to the perceptual similarity of the 'accent' such that the male

and female generalization talkers both had similar perceptual ranges for the effected segments, largely driving generalization behavior (see Reinisch & Holt, 2014). As of yet there is not a definitive and clear distinction between acoustic and perceptual similarity of talkers and how strongly they are linked to socio-indexical inferences. Furthermore, given the asymmetry observed by Kataoka and Koo (2017), it's unclear whether individual vowel categories shifting (as opposed to the entire system) demonstrate the same degree of cross-talker generalization.

Overall, theoretical accounts of perceptual learning in ideal adapter models posit that listeners draw on socio-indexical structure as a potential cause for experienced variability to guide perceptual learning behavior (e.g., Kleinschmidt, 2019; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). However, there is little work examining perceptual learning and generalization of vowels overall, and it's unclear whether individual vowel categories produce asymmetric listener behaviors that align with different forms of socio-indexical relationships. This chapter presents an experiment that specifically asks whether two vowel categories signifying different socio-indexical associations (as derived from Chapter 4) result in asymmetrical perceptual learning and generalization behavior. I hypothesize that shifts in /eɪ/, a dialect-informative category, will demonstrate learning and generalization from listeners. On the other hand, shifts in /ʊ/, a talker-informative category, will be more likely to elicit reduced learning and no generalization.

1.2 The Vowel Categories

I examine the relationship between vowel categories and social structure by comparing perceptual learning behavior across two vowel categories, /eɪ/ and /ʊ/. These two categories were selected because they exemplified differing social structure across two dimensions of interest in the earlier corpus-based studies (Chapters 4-5): dialect-informative (/eɪ/) and talker-informative (/ʊ/). In this section I will briefly overview the properties of these two vowel categories from the analyses presented in Chapters 4-5, before moving on to the experimental methods. The chosen categories represent distinct dimensions of socio-indexical structure while maintaining similarity across other vocalic properties I explored in this dissertation as a control. While there are no two vowel categories that will be perfectly balanced, these two categories approximate the best data driven solution from the findings of the corpus analyses.

As already noted, the /ʊ/ category demonstrates greater talker-specific behavior. This is based on two key findings from the corpus analysis in Chapter 4: 1) it ranks high in talker informativity but not dialectal informativity; and 2) talkers demonstrate higher divergence from their dialect area for this vowel category compared to other vowel categories. This fact suggests that talkers are not generally homogenous within their dialect areas in their productions of /ʊ/. Contrastingly, the category /eɪ/ demonstrates greater group conditioned behavior, based on two key findings from the corpus analysis: 1) it ranks high in dialect level informativity; and 2) talkers demonstrate low divergence from their dialect area for this vowel category compared to other vowel categories. This pattern suggests that talkers are generally similar to one another within their dialect areas in how they produce their /eɪ/ category. Overall, these results suggest that /eɪ/ has a greater likelihood of being informative of dialect-level variation to listeners, while /ʊ/ generally has a low likelihood of being informative of dialect-level variation. As such, I anticipate listeners will draw on experience with this relationship when processing novel variation and infer either dialectal (/eɪ/) or talker-specific (/ʊ/) causes to the variation, resulting in an asymmetry in adaptation and generalization behavior (as discussed above).

Aside from the critical distinction in socio-indexical structure, both categories demonstrate similarity in terms of their F1 and F2 overlap with other vowel categories and their cue specific variability (see Chapter 5). First, the overlap characteristics of the two vowels demonstrate generally a low degree of overlap with one another (Pillai = 0.66 on average for individual talkers). In addition, both categories share similar overlap properties with other vowels, such that they share a similar range of overlap with their tense/lax counterparts. For example, /eɪ/ overlaps with /ɛ/ and /ʊ/ overlaps with /u/ to a large degree. However, there is a slight asymmetry in their overlap properties (conditioned on individual talkers) such that /ʊ/ has a stronger degree of overlap with /eɪ/ neighbors, /ɛ/ (Pillai = 0.45) and /ɪ/ (Pillai = 0.63), than /eɪ/ does with the /ʊ/ neighbors, /u/ (Pillai = 0.66) and /o/ (Pillai = 0.79). In other words, /ʊ/ tends to show more overlap within the front of the vowel space than /eɪ/ shows within the back of the vowel space. Additionally, both categories demonstrate a similar degree of variability along F2, the primary axis for manipulation in this experiment, with /ʊ/ showing slightly greater variability in F2 than /eɪ/. Overall, the two categories are broadly matched in terms of different aspects of variability as outlined in Chapter 5. Thus, we shouldn't expect that asymmetries in perceptual

learning are the direct result of listeners expecting a priori greater overlap for one category more than the other, since on average there is reasonable separation between the categories. In addition, there's less likelihood that smaller phonetic changes will fall within a familiar range and not warrant an update (see Cutler, 2012, Babel et al., 2019). Furthermore, we shouldn't expect that sensitivity to the phonological front/back feature specification should drive any asymmetry, since both categories would be encroaching on the opposite phonological subsystem (that is, if it's purely the separation of these dimensions that matter). Additionally, an asymmetry shouldn't be driven by the general expectation of F2 variability more in one category than another. To summarize, these two vowel categories were primarily chosen for their asymmetry in socio-indexical conditioning in production and because they are relatively similar in terms of their distributional properties. Thus, all of the aforementioned confounds can predominately be ruled out as mechanisms for any perceptual learning and generalization results.

2 Predictions

2.1 Learning

Given the background provided above and the vowel categories of interest, I present the following predictions. For both conditions I expect to see learning due to listeners' a priori beliefs that vowel variability is largely thought to reflect characteristics of the talker's identity, regardless of whether it is caused by their dialect background (/eɪ/), or idiosyncratic causes (/ʊ/). Learning in this context specifically refers to an observed change in listener behavior from pre-test to post-test in the direction of exposure. Learning is predicted to occur as a targeted shift near the category boundary rather than an overall increase or decrease in categorization across steps of the continua. That is, there should be a significant interaction between the test (pre vs. post) and step of the continua. In the /eɪ/-Biased condition, listeners should categorize more items as /eɪ/ from pre-test to post-test near the categorization boundary. Whereas in the /ʊ/-Biased condition, listeners should categorize more items as /ʊ/ from pre-test to post-test near the categorization boundary. As noted in Section 1, I also predict that there may be a difference in magnitude of learning between the two conditions, such that there will be a larger magnitude of shift for the /ʊ/-Biased condition compared to the /eɪ/-Biased condition.

2.2 Generalization

The two exposure conditions should vary in terms of generalization behavior, such that we expect that because /eɪ/ is a group-informative category, listeners may likely generalize any learned pattern more flexibly because they expect it to be a pattern of multiple speakers. Because/eɪ/ varies as a function of heterogenous groups of talkers (e.g., multiple genders), we might further expect that generalization will not be constrained to same gender pairs but will extend to the other gender speaker as well. On the other hand, since /ʊ/ tends to show little to no group-informative behavior, listeners may be more likely to infer that a single talker's pattern is idiosyncratic and be unlikely to extend it to other talkers, demonstrating speaker-specific learning. Learning and generalization predictions are summarized in Table 6.1

Table 6.1 Summary of predictions for learning and generalization for each condition

Vowel	**Prediction**	**Learning**	**Generalization (Same Gender)**	**Generalization (Different Gender)**
/eɪ/	Learning & generalization	Yes: increase in /eɪ/ responses	Yes	Yes
/ʊ/	Talker-specific learning	Yes: increase in /ʊ/ responses	No	No

3 Design

The experiment consisted of three phases: a categorization pre-test, an exposure block consisting of a lexical decision task, and a categorization post-test. Participants were assigned to one of four groups: one of two exposure conditions, with one of two post-test talkers (2x2 between subjects, see Figure 6.1). The pre-test was the same across all conditions, the exposure block varied by condition, and the post-test varied within condition. The categorization pre-test consisted of two 7-step continua from /eɪ/-/ʊ/ (shake → shook; bake → book) blocked by continuum, with each step repeated 6 times and randomized within continuum block. In the

exposure task listeners were exposed to productions of words where the stressed vowel of the critical items was modified to be ambiguous between /eɪ/ and /ʊ/. The exposure conditions varied by whether the target vowel of interest was dialect-informative (/eɪ/-Biased condition) or talker-informative (/ʊ/-Biased condition). In the /eɪ/-Biased condition, the critical words contain an /eɪ/ that has been modified to shift towards /ʊ/ in word medial position, with no /ʊ/ minimal pair neighbor (e.g., *rainbow*). In the /ʊ/-Biased condition, the critical words contain an /ʊ/ that has been modified to shift towards /eɪ/ in word medial position, with no /eɪ/ minimal pair neighbor (e.g., *rookie*).

In the categorization post-test, participants were given the same categorization task from the pre-test (same talker, same items), followed by a categorization task from a novel talker using the same lexical items as continua. Within each condition, listeners were assigned to one of two novel talkers for the post-test categorization task: a novel female talker (T2_F), or a novel male talker (T3_M). The same gender and different gender pairings aimed at assessing whether cross-talker generalization was restricted to same gender speakers or would extend to different gender pairs. Additional details of the main experimental procedure will follow the description of the materials, norming, and stimuli synthesis methodology (see Section 4-6 for stimuli and norming details and Section 7 for more details about the main experiment).

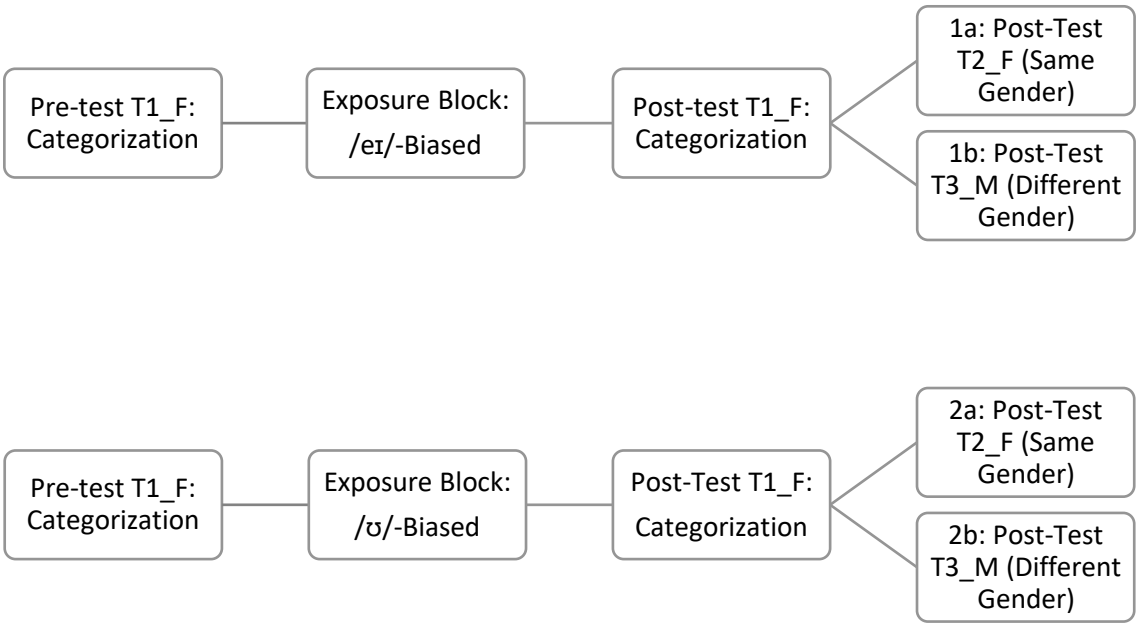

Figure 6.1 Illustration of the experimental design. The top panel represents Condition 1a and 1b; the bottom panel represents Condition 2a and 2b.

4 Materials

Items for the lexical decision task included 160 multisyllabic items per condition, composed of 40 critical items (20 unique critical items for each condition), 60 filler items, and 80 phototactically licit non-words (see Appendix A for full set of critical items, N = 180). Of the 40 critical items, 20 contained the target vowel /eɪ/, and 20 contained the target vowel /ʊ/. The critical items (N = 40) were chosen based on criteria including the position of the critical vowel, the presence of tense/lax counterparts to the critical vowel, and frequency of the word. All critical items contained the critical vowel in the primary stress position of multisyllabic utterances and did not contain a minimal pair with the /eɪ/ or /ʊ/ counterpart. Additionally, critical items were selected for the absence of their tense/lax counterparts in other positions of the word. The critical items were further inspected and removed if there was a risk of critical items being perceived as other near minimal pairs along the continuum from /eɪ/ to /ʊ/ (e.g., *skater* → sk/ʊ/ter was not used due to perceptual similarity to *scooter*). In order to prevent

participants receiving information about the entire vowel space, filler words (N = 60) and non-words (N = 80) were selected based on the following criteria: /æ/ or /a/ in initial stressed position and did not include the critical target vowels or their tense/lax counterparts anywhere else in the word. Non-words were created by using real word items and swapping phones to create phototactically licit non-words sharing similar properties to other words. The vowel categories /æ/ and /a/ were selected as vowel categories present in filler words because they generally have decreased overlap with /eɪ/ and /ʊ/, and to limit participants' exposure to low vowels, a distinct subsystem from the critical vowels. Two monosyllabic minimal pairs were selected as items for test categorization: *shake-shook, bake-book*. A full list of the stimuli used in the experiment is provided in Appendix A.

4.1 Recording

All words and non-words were recorded by three American English talkers in a sound attenuated booth using Logitech tabletop microphone with each participant given instructions to ensure similar distance from the microphone. Each speaker was given training to elicit productions as similar as possible. All critical items for the exposure task were recorded as real-word and non-word pairs, once normally and once with the target vowel swapped (e.g., *maple* → *m/ʊ/ple*). The stimuli list was randomized, with the critical real-word and non-word pairs occurring non-consecutively, and every word was repeated three times by the talker. Appendix B provides the complete word list from recording sessions. In addition to the word list elicitation, talkers were asked to read a short reading passage (from Fridland & Kendall, 2022, see Appendix B) to assess their vowel space at baseline (see talker analysis below Section 4.2).

4.2 Talker Analysis

A total of 10 candidate speakers were recorded from which three talkers were selected for the experiment: two cis-gendered female speakers (T1_F age 22; and T2_F age 19) and one cis-gendered male speaker (T3_M; age 21). All talkers were selected based on quality of recordings and similarity in demographic background, vowel space (as illustrated below), and experimenter assessment of acoustic similarity of the female talkers. All talkers currently reside in the Pacific Northwest (Oregon); speakers T2_F and T3_M were both born and raised in the Pacific

Northwest, and speaker T1_F was raised outside of the Pacific Northwest, living in several U.S. regions, but demonstrates speech patterns of the West. T3_M self-identified as non-Hispanic/Latinx and white, T2_F as Latinx and white, and T1_F as non-Hispanic/Latinx and Biracial. All talkers are native English speakers and self-report variable proficiencies in another language (Advanced – Beginning).

To get a better sense of the three talkers' vocalic behavior at baseline, both the wordlist and reading passage were forced aligned using the Montreal Forced Aligner (MFA; McAuliffe et al., 2017) and formant measures were extracted using the Forced Alignment and Vowel Extraction (FAVE; Rosenfelder et al., 2015) suite. Each talker's vowel space is presented in Figure 6.2 and Figure 6.3 below. As can be observed from these figures, all three talkers exhibit patterns associated with the Western U.S. including low-back vowel merger, /æ/ retraction, and /u/ fronting (e.g., Labov et al., 2006). To evaluate each talker's /eɪ/ and /ʊ/ categories for the experiment and whether their natural productions approximate a similar range of acoustic space, I measured the distance between the mean Lobanov (1979) normalized first and second formants of /eɪ/ and /ʊ/ using Euclidean distance (ED). Normalizing the formant values converts formants to equivalent scales, making them appropriate for use in ED which is sensitive to differences in scale between dimensions. For each talker the ED between their /eɪ/ and /ʊ/ vowels are near equivalent (T1_F = 1.83; T2_F = 1.86; T3_M = 1.63) suggesting the vowel categories occupy similar acoustic positions on average for each talker. While the reading passage data demonstrate some degree of lowering in /ʊ/ for talker T1_F, all of her back vowels appear to be somewhat lower than the other talkers in general, rather than /ʊ/ specifically lowering. However, the stimulus items depicted in Figure 6.3 demonstrate similar patterns to those of speaker T2_F, suggesting similar acoustic quality for the categories used for the experiment. Additionally, the word list items demonstrate similar pronunciation patterns across both real word and non-word productions, such that the natural end points of the non-word critical item pair are in line with the real-word productions of the same vowel category.

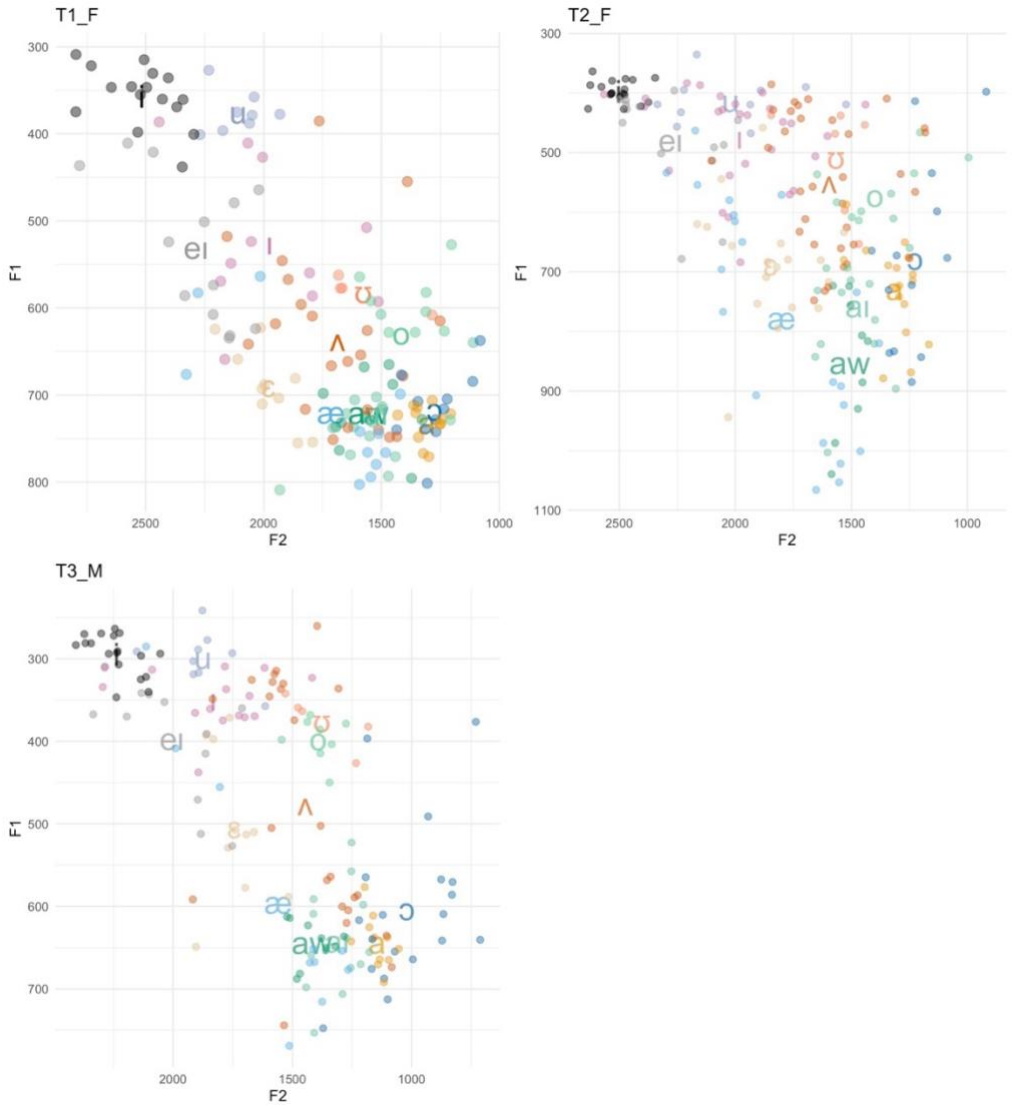

Figure 6.2 Each talker's vowel space plotted by F1 and F2 colored by vowel category. Means indicated with text, and individual tokens represented by matching color.

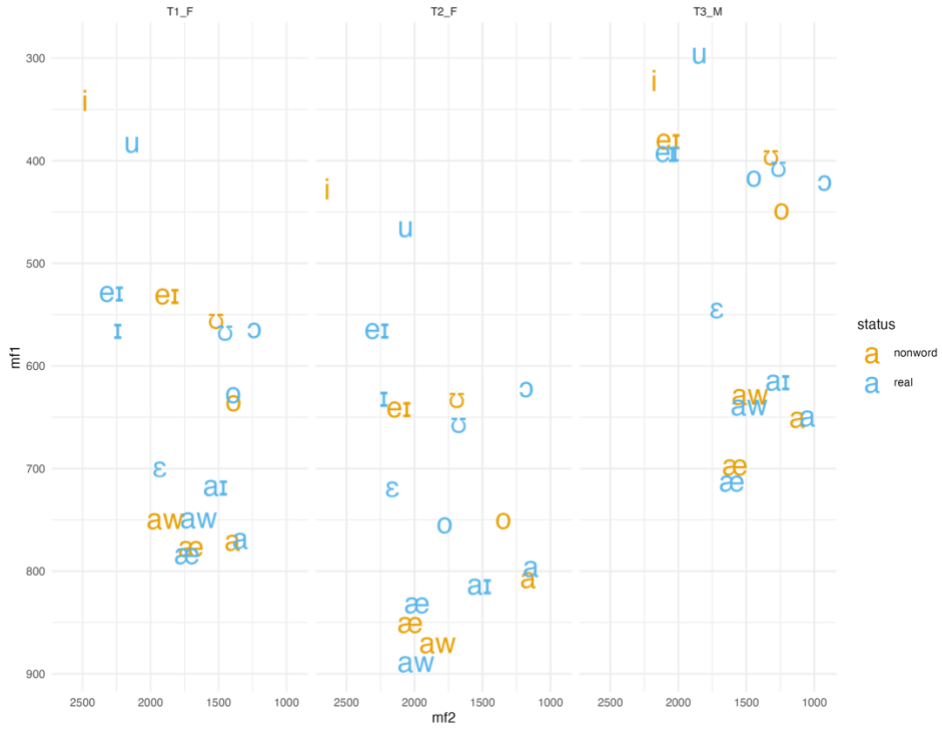

Figure 6.3 Individual means for vowel categories across real word and non-word stimulus items across all three talkers.

5 Stimuli

All categorization and lexical decision items (N = 160) were manipulated using a source-filter manipulation in Praat (Boersma & Weenink, 2018) via a custom formant continuum script (Winn, 2014). The script is designed to alter the formant structure of a single word (the 'base sound') to make it more like another word (the 'comparison sound') using LPC decomposition. For each critical item, the precursor and postcursor segments were selected from the base sound. Precursor segments were manually segmented as the preceding segment of the target vowel; the post-cursor segment(s) was/were manually segments as anything between the offset of the vowel and the end of the word (multiple phonetic segments for the multisyllabic words). Additionally, if any items had the presence of creaky voice while transitioning into the stop closure, the duration of the creaky segment was not selected as part of the target sound, the precursor, or the postcursor to avoid errors in resynthesis.

The boundaries of the target vowels were manually selected, and formant contours were visually inspected for each lexical item to ensure accurate estimates by LPC (formant settings were typically 5500hz, 5 formants for the female speakers, and 5000hz, 5 formants for the male speaker). Formant tracks were again manually corrected during script processing for each target vowel in the base and comparison sounds if needed. Each target vowel's spectral quality (F1, F2, F3) was shifted along an 11 step Bark interpolated scale between the speakers' natural end point productions. The natural end points were real word and non-word pairs produced by the speaker for the exposure task (e.g., *maple* → *muple*) and the base-sound was always the /eɪ/ word, regardless of lexical status (e.g., *beshes* → *bushes*), with the comparison-sound always the /ʊ/ word. The categorization task end points were the minimal pairs, again with the /eɪ/ word acting as the base sound and the /ʊ/ word serving as the comparison sound (e.g., *shake* → *shook*). The manipulation was set to override bandwidths and restore the original 4000Hz and a filter width of 500Hz, keeping the original upper spectrum unmanipulated for maximal naturalness. Overall, for the /eɪ/ critical items, the manipulation resulted in F1 being raised and F2 and F3 lowered, while for /ʊ/ F1 was lowered, and F2 and F3 were raised. Figure 6.5 illustrates the final categorization items for the experiment after norming (see below) layered over the talkers' vowel category means from the reading passage described in Section 4.1 above. Figure 6.5 illustrates the selected step for each item in the exposure task, layered over T1_F's vowel space from Section 4.2.

6 Norming

6.1 Participants

Participants were recruited from were recruited from Prolific (www.prolific.co)[2022] to participate in the norming tasks online. The norming experiment was programmed using PsychoPy (Peirce et al., 2022) and was integrated to the online platform using PsychoJS and Pavlovia (Peirce et al., 2022). The norming experiments included two separate experiments: the categorization task norming and the lexical decision norming. Across both experiments, each participant was subject to a headphone screening before being allowed to participate in the study (described below). All participants filled out a survey with demographic questionnaires

following the experimental norming. All participants were all first language speakers of English and reported no issues with hearing loss.

For the categorization norming, a total of 28 participants were recruited on Prolific. The participants ranged in demographic background, with 16 men and 12 women, and ranging in age from 18 – 60+, with the majority of participants ranging from 18-35, from a range of geographic regions in the U.S. (with N = 2 saying 'Other'). For the lexical decision task norming, a total of 103 participants were recruited. These participants were made up of similar demographic backgrounds, with 67 men, 35 women, and 1 non-binary participant. Participants ranged in age from 18 – 60+, with the majority between the ages of 18 – 35 (N = 63) and similarly reported a range of geographic regions in the U.S., with N = 8 reporting 'Other' as their geographic region.

6.2 Participant Headphone Screening

Participants were required to pass a binaural beats headphone screening before participation in the study. The task is drawn from Milne et al. (2020), drawing on perceptual artifacts of binaural processing. As a brief overview, when two tones of slightly different frequencies are played simultaneously in separate ears (i.e., dichotically, as through headphones) listeners perceive a third 'beat' tone of the difference between the two frequencies. For example, if a tone of 1000Hz is played in the left year and a tone of 1030 Hz is played in the right ear, listeners will perceive a third tone equal to the difference of 30 Hz as amplitude modulation. Binaural beats are only perceived for frequencies lower than 1000-1500 Hz. However, a similar phenomenon occurs in higher frequency tones when presented simultaneously (i.e., diotically, to both ears or through speakers) known as monaural beats. When listeners are simultaneously presented with beats at a high frequency range that differ by 30 Hz (e.g., 1800 Hz and 1830 Hz), they will perceive monaural beats when both tones are played non-independently to both ears. However, if the tones are split across the headphone channels (1800 Hz to the Left and 1830 Hz to the Right) listeners will perceive the two tones as one smooth tone, with no interfering beat due because of the frequency limits of binaural processing and the perceptual integration of the tones. If those same signals are heard ambiently through speakers, listeners will perceive the monaural beat. This is the essence of the headphone screen test (for a full overview see Milne et al. 2020).

Participants are presented with three pairs of tones in one trial and asked to identify the trial with the 'smoothest' tone. Two of the three pairs are designed to elicit monaural beat percepts, as the 'foils', where two tones are played differing by 30Hz within a frequency range of 1800 – 2500 Hz. The first tone of the pair will be randomly drawn from the range of 1800 – 2500 Hz and the second tone will be 30 Hz higher. The third pair is the target pair, where the tones also fall within the higher frequency range but are presented to each headphone independently. If listeners are wearing headphones, they will perceive the two foil tones as having a beat or fluctuation of the tones and the third target tone as being a smooth single tone. If they are not wearing headphones, the percepts will all be indistinguishable, and they will fail to identify the 'smoothest' tone. This process repeats for 6 trials, with stimulus items within each trial randomized. Listeners must get all 6 trials correct before being able to proceed to the main experiment and can only reattempt the screening once. All participant information above reflects participants who passed the headphone screening and completed the norming task.

6.3 Categorization Task Stimuli Norming

Each of the continua were subject to a norming task to select the final steps, a reduction from the initial 11 to the final 7 steps of the minimal pair continua for the main experiment. The norming task was implemented in PsychoPy using PsychoJS tools, and participants completed the task online with the use of headphones. Participants were presented with each step of 2 continua 6 times (11 steps x 6 repetitions x 2 continua x 3 talkers = 396 trials) and asked to identify the word they heard from the two minimal pairs presented on the screen (e.g., Press F for *shake*; Press J for *shook*). The experiment was blocked by continuum and talker, with talker order and button order counterbalanced across participants. Trials were pseudorandomized within talkers and minimal pairs such that no adjacent steps occurred consecutively, with 500ms ISI between trials and a 3000ms response time limit, at which point listeners would be alerted that no response was detected.

The proportion of real word responses across all 11 steps is presented in Figure 6.4, from which the final continuum steps were chosen. The final 7 steps were selected by choosing the cross-over point (50% /eɪ/ word responses) plus the 3 points on either side of the cross-over point. The results are displayed in Figure 6.4. For the *shake-shook* continuum, steps 1-7 were

selected for T1_F and T3_M and steps 2-7 and step 9 were selected for T2_F. For the *bake-book* continuum steps 1-7 were selected across all talkers. Figure 6.5 depicts the three talkers' original vowel plots alongside the continua used for the categorization task. In addition, it shows each talkers' categorization continua in one plot viewpoint and illustrates the similarity between distributional properties along F1 and F2 for each talker, in addition to the overall relative positions of each step in the continua. From this figure, we can see that talkers largely vary in F2 to the same degree for the continua, while for F1 the female talkers align more closely to one another and the male speaker is much more consistent in F1 cues. That is, both distributional properties and acoustic positions of the vowel stimuli for the categorization task are more similar for the two female talkers than the male is to either of the female talkers.

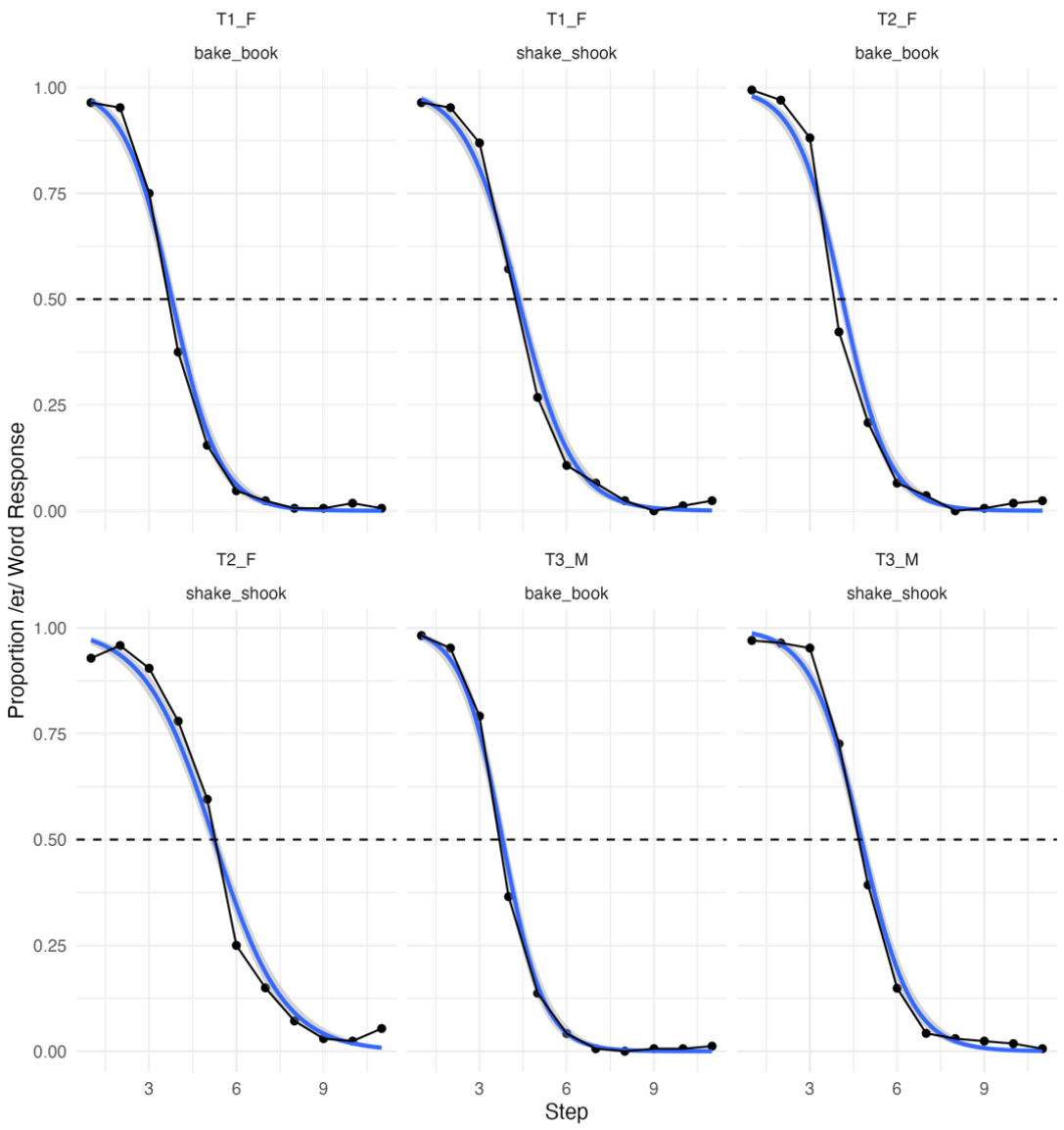

Figure 6.4 Proportion of /eɪ/ responses for categorization items in the norming task. Horizontal dashed line represents the cross-over boundary (50% /eɪ/-word response rate). Dots are averaged /eɪ/ word response across subjects, and the blue line is a binomial model of the responses. Step 1 represents the /eɪ/ end point and step 11 represents the /ʊ/ end point.

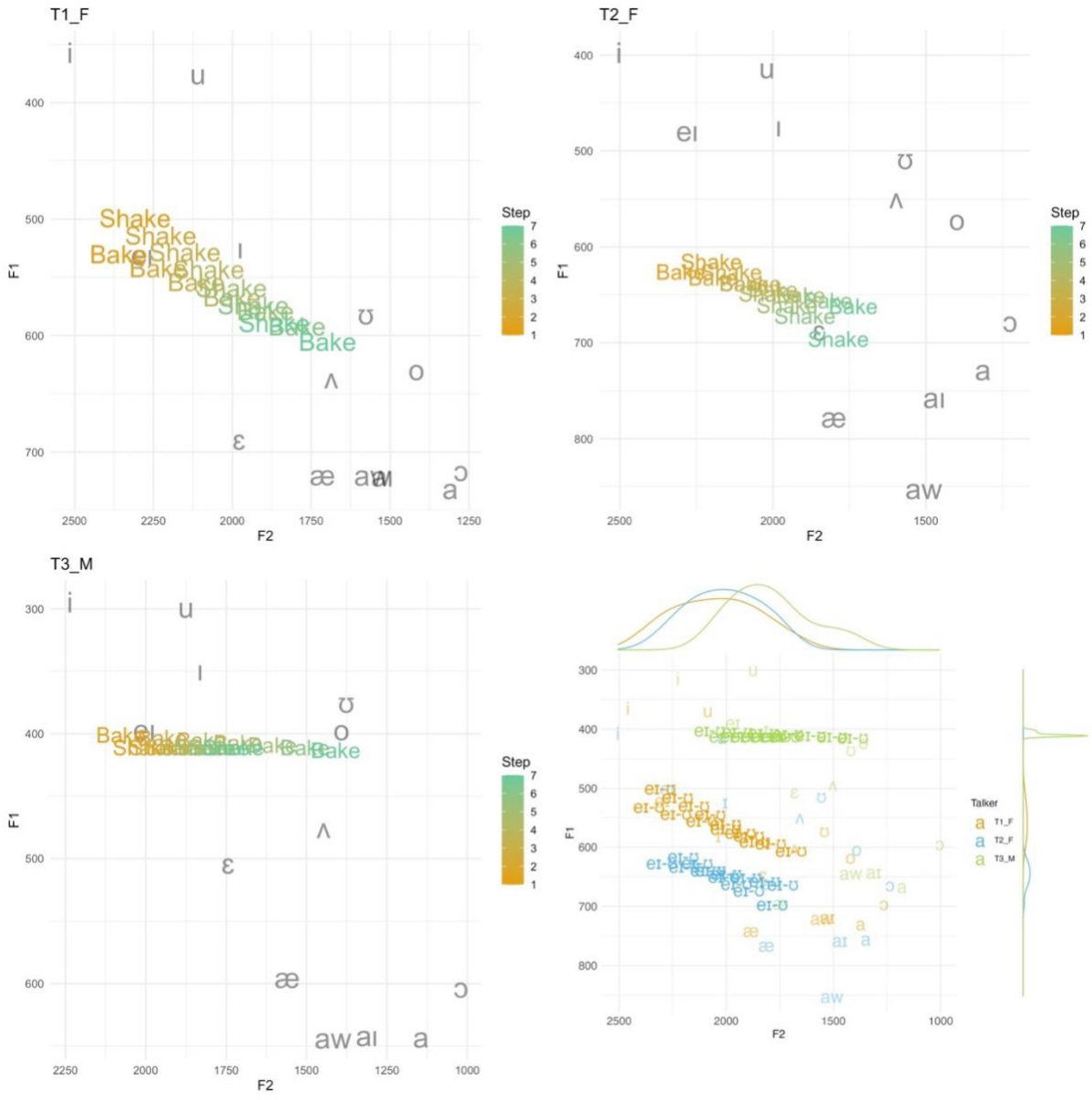

Figure 6.5 The three talkers' final continua overlaid on their average raw unsynthesized vowel space (grey). Final figure represents all three speakers' continua and raw vowel spaces on one plot with the distributional properties of their continua appearing as ridges along the X and Y axes.

6.4 Lexical Decision Task Exposure Stimuli Norming

To determine which step of the continuum to use for each lexical item in the exposure task (i.e., the step that was maximally ambiguous), participants (N = 103) recruited were from Prolific (www.prolific.co)[June 2022] and completed a lexical decision task online on the

exposure continua. Participants were randomly placed into one of 4 conditions consisting of 10 word/non-word continua (1a = 27; 1b = 22; 2a = 24; 2b = 30). Listeners were exposed to only one vowel category throughout the task (e.g., only /eɪ/ words in condition 1-2) blocked by lexical item (e.g., *pastry* → *pustry*). Participants were presented with each step of the 11-step exposure continuum, responding with either 'word' or 'non-word', with each step repeated 6 times, totaling 660 trials. Trials were pseudorandomized within lexical blocks with no adjacent steps occurring consecutively and a 500ms ISI between trials and 3000ms response time limit. Following this initial norming, some lexical items were subjected to resynthesis and a new round of norming with participants recruited from Prolific with a similar demographic make-up as the previous participant pools (3a = 19). In Figure 6.6 below, there are duplicated lexical items as a result. Figure 6.6 illustrates the acoustic range of the 11 steps for each of the lexical items in the norming experiment. As illustrated, the items vary widely in acoustic realizations of their start and end points, highlighting the necessity for norming to identify the most ambiguous point.

The proportion of word endorsements at each step of the continuum was calculated and the most ambiguous step was chosen by selecting the step that approached 50% word endorsement rates. Figure 6.7 shows the average proportion of real word responses for each step and word in the experiment, with a binomial response curve overlaid. As noted above, some items in initial norming did not receive real word response curves as expected, with some items' steps never dropping below the 50% point, or steps consistently hovering near the 50% response rate. Those items were subject to resynthesis and renorming, followed by a selection of the best choice from the two options. Table 6.2 shows which steps (and item if subject to renorming) were chosen along with the proportion of word endorsement rates during norming. The average step chosen was 5 for the /eɪ/-Biased condition and 4 for the /ʊ/-Biased condition.

To illustrate the acoustic properties of the exposure items, all tokens were force aligned using MFA (McAuliffe et al., 2017) and formant values were extracted using FAVE (Rosenfelder et al., 2015). Figure 6.8 illustrates the vowel space properties of the three phases of the experiment in a composite image for the exposure talker. In the first graph of the image, the talker's final categorization continua for pre and post-test are overlaid on the exposure talker's original raw vowel space for comparison. The middle figure illustrates the final critical items in exposure, colored by the condition, and the final figure overlays both the categorization stimuli

and the exposure stimuli. Taken together these illustrate the distributional properties of the stimuli, whereby the critical items for both conditions demonstrate similar distributional properties with the mean and standard deviation similar across F1 and F2. Similarly, the categorization continua span a wider range of F1 and F2, and with F1 demonstrating an overall higher mean F1 (i.e., lower F1) compared to the exposure stimuli.

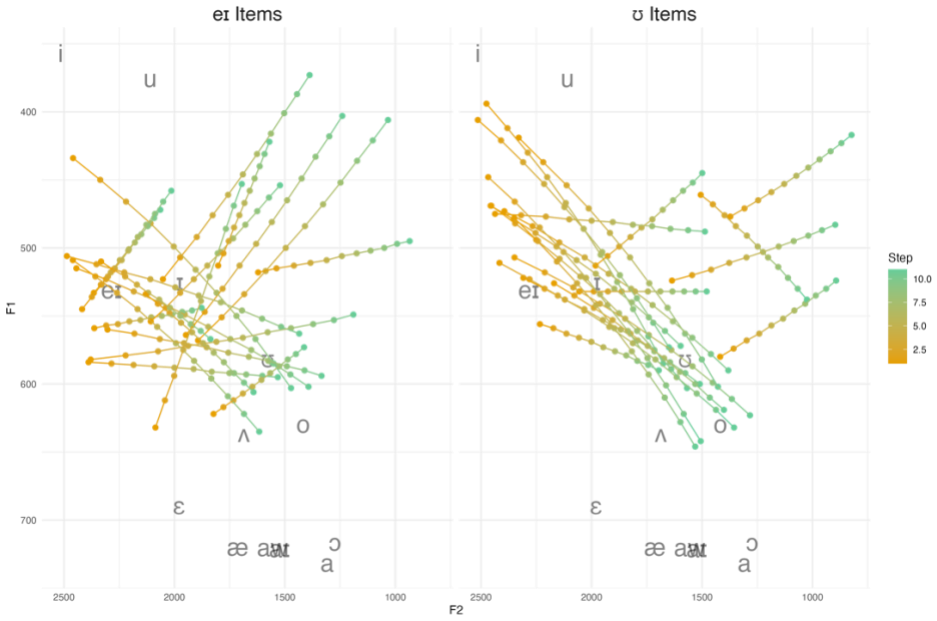

Figure 6.6 Real word – Non-word continua 11 steps connected by a line, overlaid on top of T1_F's vowel space for both /eɪ/ items and /ʊ/ items from the reading passage data for comparison.

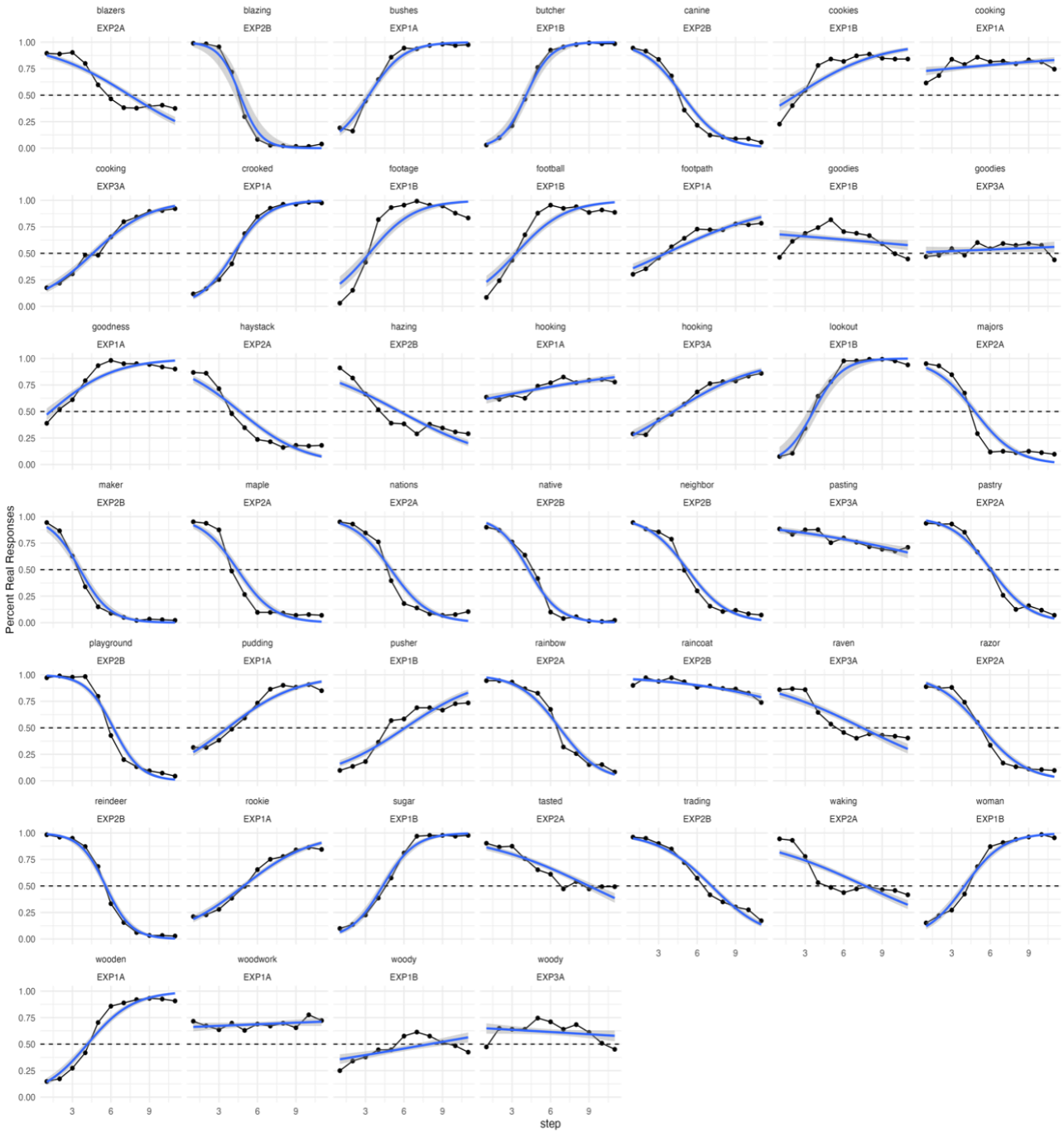

Figure 6.7 Proportion of word-responses for exposure words. Horizontal dashed line represents the selection criteria (50% word response rate). Dots are averaged real word response across subjects, and the blue line is a binomial model of the responses. Step 1 represents the /eɪ/ end point and step 11 represents the /ʊ/ end point.

Table 6.2 Step chosen for each lexical decision exposure task, with proportion of 'real word' responses from norming. Average step and proportion for each condition listed at the bottom of the table.

/eɪ/-Bias Condition			/ʊ/-Bias Condition		
Stimuli	**Step**	**Proportion Word Response**	**Stimuli**	**Step**	**Proportion Word Response**
blazers	5	0.56	bushes	3	0.46
blazing	5	0.35	butcher	4	0.47
canine	5	0.40	cookies	3	0.54
haystack	4	0.49	cooking (3A)	5	0.46
hazing	4	0.52	crooked	4	0.44
majors	4	0.61	footage	3	0.45
maker	4	0.39	football	3	0.46
maple	4	0.48	footpath	4	0.52
nations	5	0.40	goodies (1B)	1	0.45
native	5	0.44	hooking (3A)	8	0.47
neighbor	5	0.50	lookout	3	0.38
pastry	6	0.47	pudding	4	0.47
playground	6	0.43	pusher	4	0.40
rainbow	7	0.37	rookie	5	0.46
raven	5	0.52	sugar	5	0.53
razor	5	0.55	woman	4	0.45
reindeer	6	0.37	wooden	4	0.45
tasted	7	0.48	woodwork	7	0.55
trading	6	0.56	woody (3A)	10	0.54
waking	5	0.48	goodness	2	0.50
Average	0.15	0.47		4.3	0.47

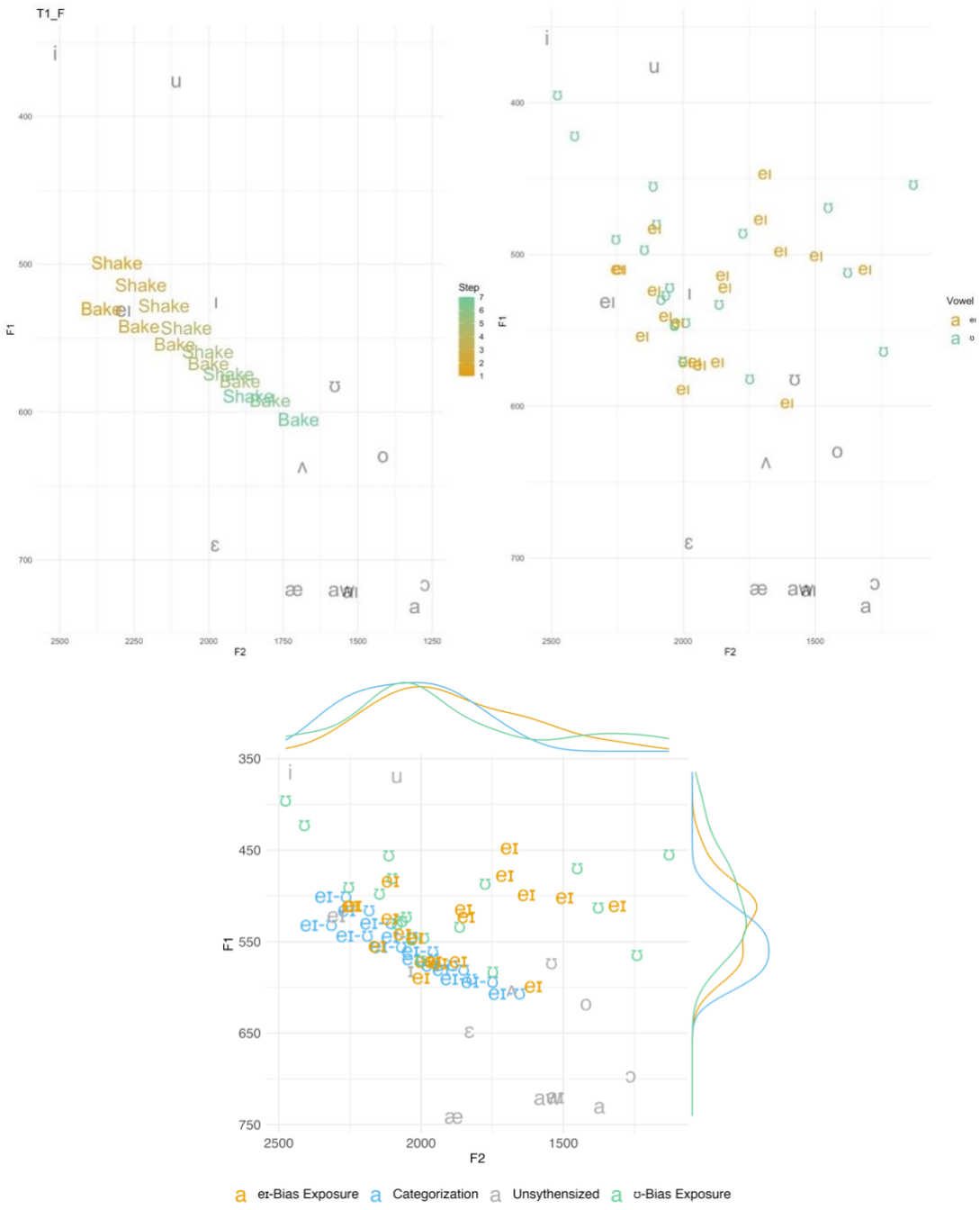

Figure 6.8 Position of the stimuli in the vowel space overlapped with the talkers' original point vowels for reference. Final plot shows exposure and categorization items over T1_F's original vowel space with distributions of F1 and F2 for each category.

7 Main Experiment

7.1 Procedure

Following the norming task, the main experiment was implemented in PsychoPy using PsychoJS tools and integration with Pavlovia. All participants (see Section 7.2 below) took the same pre-test categorization task, identifying minimal pairs along a seven step continuum from /eɪ/→/ʊ/ from talker T1_F (see Section 5 for details about stimuli creation and step selection). In each trial participants heard an auditory stimulus and were asked to categorize it as one of two words, differing only in the vowel (e.g., *shake* and *shook*). Listeners were presented with the key to press and item indicated on the screen (e.g., "F" + "*shake"* on one side of the screen and "J" + "*shook"* on the other), with button and item matches counterbalanced across participants. Participants were given two unique minimal pair continua for the categorization task, each with 7 steps, repeated 6 times each, for a total of 84 trials. The task was blocked by minimal pair, and steps were pseudo-randomized such that no adjacent steps occurred consecutively, and the same step would not be repeated sequentially across trials. After each trial a blank screen was displayed for 500ms followed again by the response screen indicating which options to select. Participants were given 3000ms to respond, at which point they would be told no response was detected. After 50 trials participants were given the opportunity to take a break.

Following the pre-test categorization phase, participants were randomly assigned to one of two exposure blocks with same talker (T1_F): the /eɪ/-Biased condition or the /ʊ/-Biased condition. The exposure block was a lexical decision task, where participants were presented with an auditory stimulus and then indicated whether what they heard was a word or a non-word. Similar to the categorization task participants were given 3000ms to respond on each trial, at which point a message would appear that indicated no response was detected, and the experiment proceeded to the next lexical item. In the /eɪ/-Biased condition participants were exposed to lexical items where the /eɪ/ target vowel is shifted to an ambiguous point between /eɪ/ and /ʊ/. In the /ʊ/-Biased condition participants were exposed to lexical items where the /ʊ/ target vowel is shifted to an ambiguous point between /eɪ/ and /ʊ/. Ambiguous items were selected through a pre-test lexical decision task, using the item that participants responded to as a word 50% of the time (see Section 6 above for further details).

Following the exposure block, participants completed a post-test, which was the same categorization task as the pre-test, using the same talker as in the pre-test and exposure (T1_F). Then, participants were asked to complete another categorization post-test block on a speaker they had not previously heard. Participants were randomly assigned to either a novel talker of the same gender (T2_F), or a novel talker of a different gender (T3_M). For both post-test blocks, the categorization task mirrored the properties of the pre-test categorization task: stimuli were two minimal pair continua (shake → shook; bake → book) each with 7 steps repeated 6 times, with the steps selected from an initial categorization norming task.

7.2 Participants

Participants were recruited from Prolific (www.prolific.co)[2022] and participated in the experiment online. Participants were required to be located in the U.S. and speak English as a first language to enroll in the study. All participants were paid for their participation at an hourly rate of $10/hour allocated across the estimated time to complete the experiment. Participants were required to pass the same headphone screening as described for the norming task before being able to complete the experimental study. A total of 210 participants completed the experiment. While Prolific enrollment required participants to be L1 speakers of English and be from the U.S., occasionally participants made it through initial screening into the experiment and nonetheless reported non-native speaker status in the screening questionnaire at the end of the experiment. Thus, prior to analysis, the data were processed to remove participants who reported being non-native speakers of English (N = 2) or participants who reported having hearing difficulties (N = 6). Additionally, participants who performed with less than 85% accuracy on filler items in the lexical decision task were removed from the pool (N = 1), though still paid for their time. After removing participants, a total of 201 participants are considered for analysis.

The participants were placed randomly into one of four conditions (see Section 3), /eɪ/-Biased with the same-gender generalization talker (1a, N = 56)), /eɪ/-Biased other-gender generalization talker (1b, N = 45), /uh/-Biased same-gender (2a, N = 54), and /ʊ/-Biased other-gender (2b, N = 48). The participants were distributed across genders according to participants self-report: female (N = 72), male (N = 124), non-binary (N = 5), and two who declined to report. Participants were distributed across age ranges, with the majority ranging from 18 – 35

years old (N = 138), followed by 36 – 60 years old (N = 59), 60+ years old (N = 7), and 6 participants who declined to answer. The majority of participants' self-reported racial identity was white (N = 105), followed by electing not to respond (N = 70). The remaining participants reported being multi-racial (N = 9), Asian/Asian-American (N=10), Black/African American (N = 9), American Indian (N = 1), Jewish (N = 1), and Hispanic/Latinx (N = 5).

Participants' regional backgrounds across the U.S. were mixed: South (N = 59), West (N = 44), Midwest (N = 37), Northeast (N = 35), Southwest (N = 10), North (N = 5), and some who preferred not to answer (N = 20). I do not expect regional background to be predictive of perceptual learning or generalization behavior, as the vowel categories under exploration are examined due to the fact that they are either regularly conditioned on dialect factors (/eɪ/) or regularly conditioned on talkers (/ʊ/) and the directions of shift do not occur within any regional dialects (that I'm aware of). While participants' boundaries may be relative to their dialect areas, the overall effect of learning and generalization should not be impacted by dialect-specific expectations of variability based on the corpus analyses in Chapters 4-5. Thus, I do not look at regional background in any of the analyses reported in this chapter.

8 Analysis & Results

Any trials where the reaction time was greater than 2.5s (including no-response time-outs) were removed from evaluation (N = 70 trials < 1% of data). Because the data were collected online, participants' experimental setups were variably sensitive in the way reaction time was recorded, so 2.5s was chosen as a practical cutoff for further data analysis based on the distribution of the data. Given these data, in the sections below I will analyze the effects of each phase of the experiment independently, first looking at the lexical decision exposure (Section 8.1), followed by the learning results (Section 8.2), and ending with the generalization results (Section 8.3).

8.1 Exposure: Lexical Decision Task

Figure 6.9 illustrates the lexical decision task endorsement rates across participants by condition. Here we see that, as expected, the filler real word items receive real word endorsements 91% of the time across both conditions, and filler non-words receive real word

endorsements 17% (/eɪ/-Biased) and 19% (/ʊ/-Biased). Overall listeners show high accuracy for identifying real words among the filler items. On the other hand, the critical items across the two conditions show lower rates of endorsement, with critical items in the /ʊ/-Biased receiving real word endorsements only 50% of the time and the /eɪ/-Biased condition 63%. Looking at Figure 6.10, we see that both conditions show an increase in word endorsement rates throughout the experiment, as trial number increases. However, the /ʊ/-Biased condition only moderately increased, reaching a max of 54% real word responses in the last 27 trials. On the other hand, the /eɪ/-Biased condition reaches a max of 70% real word responses in the same span. Overall, this suggests listeners had difficulty identifying critical items as real words in both conditions, but more so in the /ʊ/-biased condition. In the next section I will turn to examine whether the low rates of real word responses influence the patterns listeners learned.

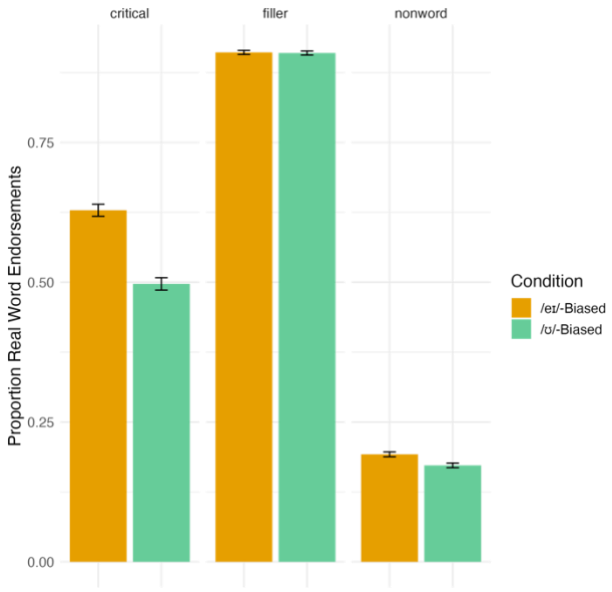

Figure 6.9 Lexical decision exposure responses. Y axis represents the average proportion of real word endorsements across participants, colored by condition, and faceted by stimulus type. Error bars represent standard error.

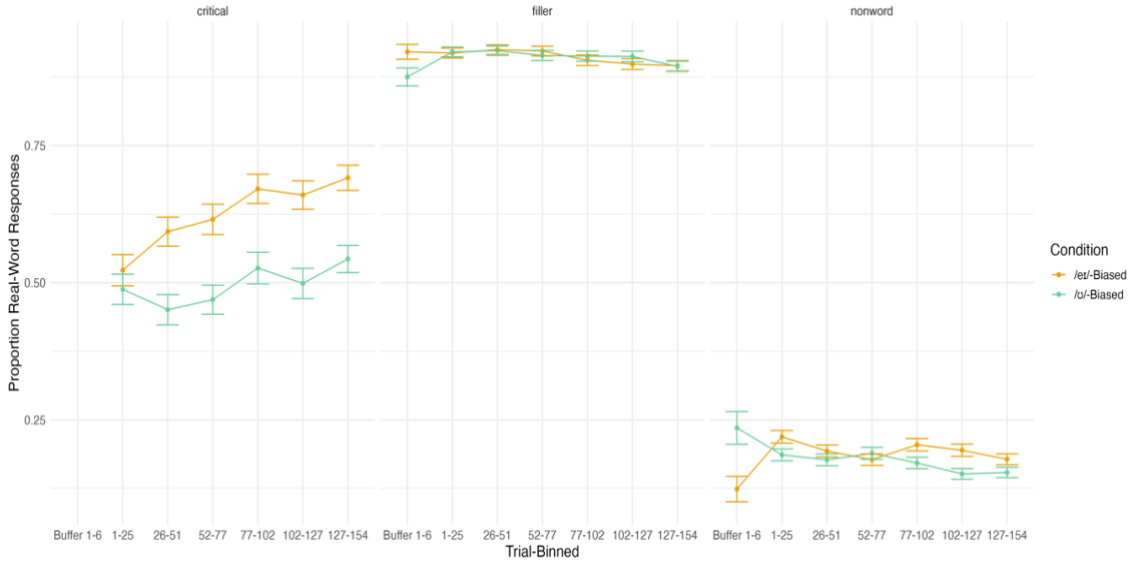

Figure 6.10 Rates of real word responses across binned trials faceted by stimulus type and colored by condition. Error bars represent standard error.

8.2 Learning

Turning to examine the learning effects, Figure 6.11 below illustrates the raw categorization data from pre- to post-test across the two conditions. The Figure illustrates proportions of /eɪ/ responses and the standard error across each step of the continuum for each condition at pre- and post-test. Here listeners exhibit similar response patterns at pre-test, albeit with an identifiable bias towards /ʊ/. Additionally, these data illustrates a trend towards learning in the /ʊ/-Biased condition, where listeners show increased /ʊ/ responses after exposure (post-test). This suggests that despite the low rates of word endorsement during exposure, listeners did learn the /ʊ/→/eɪ/ shift. However, listeners in the /eɪ/-Biased condition show a *decrease* in /eɪ/ responses after exposure (i.e., the opposite of the exposure pattern), despite having higher rates of word endorsements to critical items during exposure. The pattern of the /eɪ/-Biased condition shows that listeners did change behavior from pre- to post-test but did not learn the /eɪ/-shifted pattern from exposure. Rather, unexpectedly, the results illustrate a change in their behavior similar to that of the /ʊ/-Biased condition. To evaluate the credibility of these effects, I first

analyze all of the data together using Bayesian logistic regression. Then I turn to examine each condition independently to gain more insights into the observed patterns.

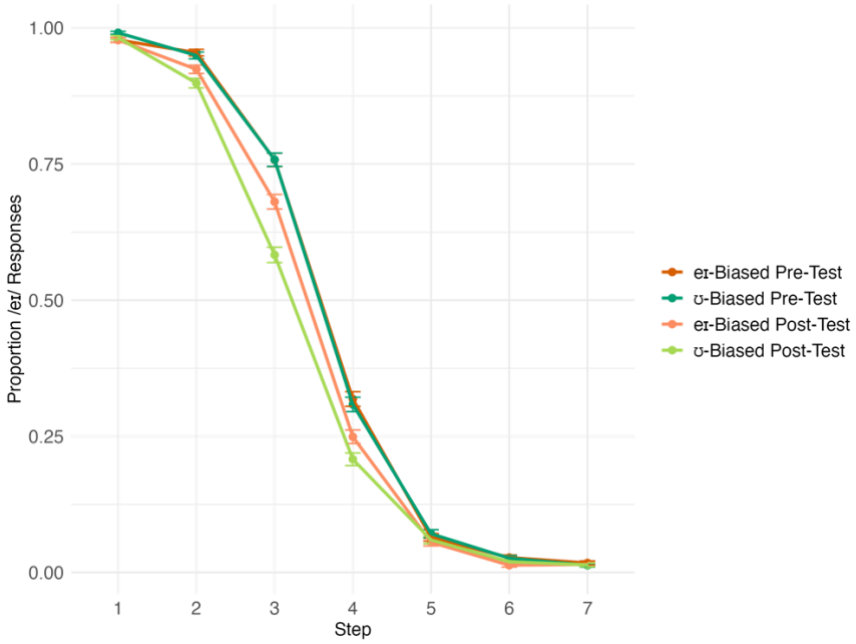

Figure 6.11 Raw learning results plotted with step as a categorical factor for ease of viewing. Error bars represent standard error.

The analyses below use Bayesian logistic regression using the brms package (Bürkner, 2017) in R (R Core Team, 2021) to draw inferences about the credibility of the effects across the two conditions. Based on the hypothesized learning effects in relation to the observed effects between conditions (as illustrated in Figure 6.11), Bayesian modeling is an appropriate choice to further understand the effect of learning across the two categories. Bayesian models are increasingly being used over more traditional null hypothesis significance testing due to their flexibility in fitting complex random effects structures and interpretability of effects (see Vasishth et al., 2018 for a tutorial in phonetic sciences). One key advantage of using Bayesian statistics for the study in this chapter is that we can focus our attention on quantifying the uncertainty around the magnitude and direction of the effects by identifying a credible interval of plausible values without reference to an unobserved hypothetical distribution used in traditional null hypothesis frameworks. For the data presented here, these benefits offer insights into whether the same underlying model generated the data; that is do the two conditions represent the same underlying effect of different magnitudes or do they represent different underlying

effects. Integrating prior domain knowledge alongside Bayesian inference will overall provide a more nuanced perspective of listener behavior.

8.2.1 Analysis: Overall Model

To begin, I briefly consider whether there is a statistical difference between the two conditions' effects observed in Figure 6.11. The change in listener behavior from pre to post-test in the /eɪ/-Biased condition is surprising and did not follow the shift in exposure. In this section I use Bayesian modeling to ascertain to what extent the patterns in the /eɪ/-Biased condition are the same as the patterns in the /ʊ/-Biased condition. That is, is the decrease in /eɪ/ responses credibly different between the two conditions, or do they represent an overall similar change in listener behavior from pre-test to post-test. In this section I aim to specifically evaluate whether there is a distinction between the two conditions, using Bayesian model comparisons, rather than evaluating the size and credibility of the effect. As such, this section primarily aims to evaluate whether there is a difference between them, after which I go on to evaluate more critically each individual condition including the magnitude of changes.

To evaluate the effect of exposure condition on the two categories above, I compared two models. The first model is an initial model with a subset of fixed effects and their interactions (Step, Test) and random slopes for the interaction of step and test by participant, and minimal pair by participant. The second model is a reduced form of model 1, where the effect of condition has been removed. Both models are fit using the brms package (Bürkner, 2017) in R, with 4 chains and 8000 iterations per chain, of which half are warmup iterations (4000). All fixed effect and random effect parameters were fit with weakly informative and regularizing priors (*Normal*, $\mu = 0$, $\sigma = 1$) and the correlation terms were specified with the default LKJ prior. Models were inspected visually and through model diagnostics to evaluate convergence; chains have converged ($\hat{R} = 1.0$ for both models) and there is no evidence of high collinearity between predictors in either model. The reduced model and the full model are then compared to derive a Bayes Factor, using the bayestestR package (Makowski et al., 2019). Bayes factors are indicative of the relative evidence of one model over another and, similarly, at the predictor level evaluates the parameter in relation to a specified null value. Bayes factors are generally related to the classical interpretation of a p-value, testing how unlikely the data are if the null is true. For

model comparison, a Bayes Factor is interpreted as evidence for or against the null hypothesis (specified by the model) whereas for parameters in can be seen as evidence for or against a null effect. As a general rule of thumb, if the Bayes factor is greater than 1, it can be interpreted as evidence against the null, and a Bayes factor greater than 3 can be considered substantial evidence against the null (Raftery, 1995; Wetzels et al., 2011). Similarly, a value of smaller than 1/3 can be interpreted as evidence in favor of the null hypothesis. Here, the two models show a Bayes factor of 2.39, providing weak evidence for the alternate hypothesis and lending credibility to an overall difference between exposure conditions. That is, despite an increase in /ʊ/ responses at post-test for both conditions, the two conditions appear to be qualitatively different from one another and are not showing the exact same patterns of listener behavior. To understand the differences between conditions, I will first examine the individual effects within the full model, and then go on to examine each condition individually in Section 8.2.2.

Bayes Factors for each effect are calculated in relation to a range of parameter values that equate to a null effect set by the researcher. The range used to evaluate effects is -0.18 – 0.18, as an indication of a small effect size (Makowski et al., 2019). Thus, the Bayes Factor indicates whether the observed posterior distribution of parameter values is credibly different from these values. Table 6.3 below provides the model parameters, coefficients, and Bayes Factors for each parameter and the interpretation of the credibility of the effect (e.g., weak, strong) following rules of thumb outlined in Raftery (1995). Overall, there is very strong evidence for the effect of the intercept, where listeners show a general preference for /ʊ/ across the board (β = -0.91, BF > 150). In other words, listeners generally demonstrate more /ʊ/ responses along the continua, which may indicate a bias towards /ʊ/ percepts for the continua stimuli. There is positive evidence for the main effect of Step and Test, such that listeners show a decrease in the log-odds of /eɪ/ responses as step increases, and a decrease in the log-odds of /eɪ/ responses at post-test. There is also weak evidence for the interaction of Step and condition, such that there is a greater decrease in the effect of step in the /ʊ/-Biased condition (β = -0.44, BF = 1.17). Meaning, from pre-test to post-test listeners show a shift towards the /ʊ/ end of the continuum and the effect is of greater magnitude for the /ʊ/-Biased condition. All other interactions and the main effect of condition show weak or positive evidence against the alternate hypothesis (i.e., favoring the null hypothesis).

Altogether, these results demonstrate evidence against the learning hypotheses that listeners would learn the exposure pattern across both conditions and that /ʊ/ would demonstrate greater magnitude of learning. Rather, it appears that listeners have learned the exposure shift of the /ʊ/-Bias condition but have not learned the shift of the /eɪ/-Biased condition. There appear to be qualitative differences between the listeners' behavior between the two conditions, despite generally trending in the same direction from pre-test to post-test. To better understand the difference between the two conditions, I turn to examine models for each condition individually. For the remainder of the analyses, I will examine the posterior distributions to assess the confidence, magnitude, and direction of the effects in greater detail.

Table 6.3 Bayes factor summary for the model parameters, with interpretation of the magnitude of evidence schema following Raftery's (1995) scale. Bayes factor values ranging from 3:20 indicate positive evidence, 20:150 = Strong, > 150 = Very Strong.

Parameter	Model Coefficient	Bayes Factor	Log Bayes Factor	Evidence of Effect
Intercept	-0.91	> 150	11.33	Very Strong
$Step_{[scaled]}$	-1.83	4.98	1.55	Positive
$Condition_{[/ʊ/\text{-}Bias]}$	-0.11	0.13	-1.76	Positive (against)
$Test_{[post\text{-}test]}$	-0.37	6.60	3.23	Positive
$Step_{[scaled]}$*$Condition_{[/ʊ/\text{-}Bias]}$	-0.44	1.17	0.35	Weak
$Step_{[scaled]}$* $Test_{[post\text{-}test]}$	0.15	0.14	-1.66	Positive (against)
$Condition_{[/ʊ/\text{-}Bias]*}Test_{[post\text{-}test]}$	-0.32	0.95	0.57	Weak (against)
$Step_{[scaled]}$* $Condition_{[/ʊ/\text{-}Bias]*}Test_{[post\text{-}test]}$	0.27	0.34	-0.89	Weak (against)

8.2.2 Analysis: Individual Conditions

For describing the direction of effects of each condition more clearly, each condition's dependent variable was recoded to reflect the hypothesized learning bias in each condition. Specifically, for the /eɪ/-Biased condition, the response variable is /eɪ/ responses. For the /ʊ/-Biased condition, the response variable is /ʊ/ responses, and the steps have been reversed to indicate the greatest /ʊ/ responses on the left (step 1) and lowest /ʊ/ responses on the right (step 7), illustrated in Figure 6.12 for clarity. Based on the results above, but counter to initial

expectations, we can expect a *decrease* in /eɪ/ responses from pre-test to post-test in the /eɪ/-biased condition. On the other hand, we should expect an *increase* in /ʊ/ responses in the /ʊ/-Biased condition from pre-test to post-test.

Each condition was modeled using Bayesian logistic regression in R (R Core Team, 2021) using the brms package (Bürkner, 2017) with 4 chains and 8000 iterations per chain, with 4000 warm-up iterations (half of the total iterations). Each model followed the same specifications, with a binary dependent variable of preference (0 'dispreferred' and 1 'preferred') using the Bernoulli distribution and a logit link function. An interaction between Test and Step and their main effects were specified as predictors in the model with a weakly informative prior (*Normal* $\mu = 0$, $\sigma = 1$) applied across all main effect predictors, including the intercept. Test was treatment coded with 'Pre-Test' as the reference level; Step was treated as a numeric predictor centered on 0. The random effects structure included by item and by participant random slopes for Test, Step, and their interaction. All random effect terms were fit using regularizing priors (*Normal* $\mu = 0$, $\sigma = 1$), and correlation terms were given default LKJ priors. Both models were visually inspected, and model diagnostics were evaluated; accordingly, chains have converged ($\hat{R} = 1.0$) and there is no evidence of high collinearity between predictors in either model.

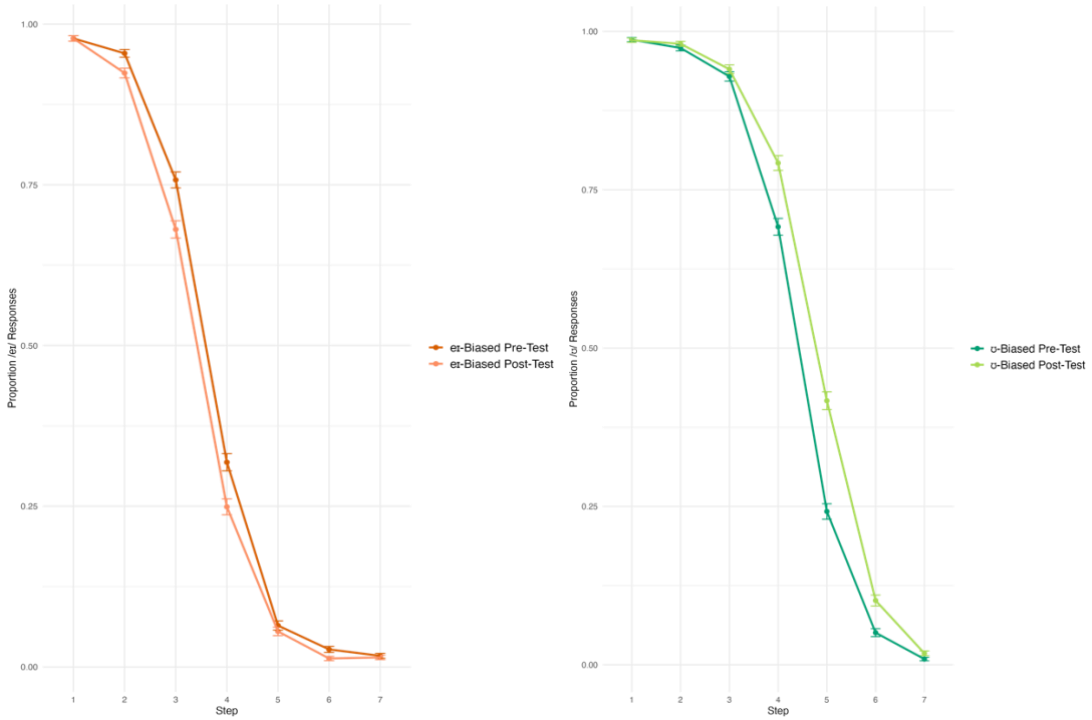

Figure 6.12 Response curves for each condition by pre- and post- test. For the /ʊ/-Biased conditions steps are reversed and /ʊ/ responses are given a value of 1, and /eɪ/ responses are given a value of 0. In the /eɪ/ biased condition, Step is in the original order, and /eɪ/ responses are given a value of 1 and /ʊ/ responses are given a value of 0. Error bars represent standard error.

8.2.2.1 /eɪ/-Biased Condition:

With the initial insights in mind, this section examines the /eɪ/-Biased condition results, with a focus on describing the model estimates posterior distributions to get a better understanding of the stability and certainty of effects in the model. First, to summarize the main effects, the /eɪ/ model validates the larger overall model above, demonstrating an overall bias towards /ʊ/ responses at Pre-Test and Post-Test, with a further decrease in /eɪ/ responses at Post-Test (β = -0.33, *SE*= 0.10, 95% CI = -0.52 – -0.14), and a weak positive effect between Test and Step (β = 0.09, *SE*= 0.10, 95% CI = -0.11 – 0.28). Figure 6.13 depicts the model's posterior estimated density functions (red fill) alongside the estimated density function of the prior (blue fill). Of note before getting into specific decisions, one can observe that the posterior estimates for the interaction term (Step*Test) and the main effect of Test have narrow distributions (e.g., low standard deviation), which suggest higher certainty around the estimated model coefficients.

On the other hand, the Step predictor has a much wider distribution and skew demonstrating lower certainty in the estimates of the model coefficients and some more extreme plausible coefficients (towards 0). The variability in Step may be driven by the variance explained by other terms in the model, including by participant slopes, which make conditional estimates of Step as a main effect more variable. Additionally, it's possible that the model estimates are somewhat unstable for the main effect of Step due to the fact that some of the participants show quasi or complete separation such that Step perfectly separates their binary response (e.g., all /eɪ/ or all /ʊ/). Overall, there is no evidence of learning of the exposure pattern, and listeners appear to be reducing their /eɪ/ responses at post-test, counter to the initial predictions and unexpected given prior literature.

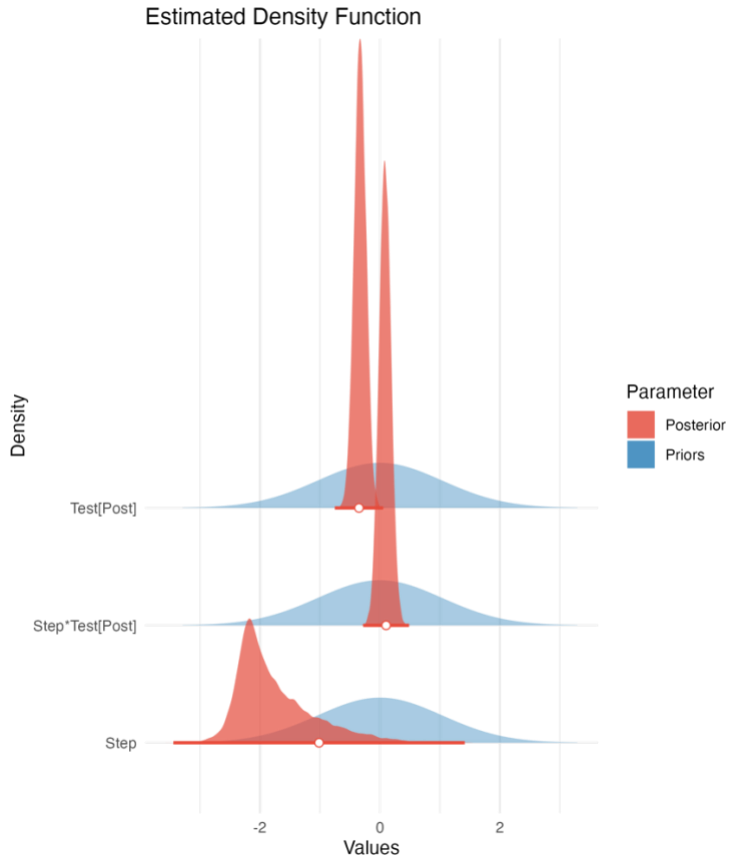

Figure 6.13 Estimated density function of posterior estimates (red) overlaid the estimated density function of the prior distribution (blue). The circle under the posterior represents the mean and the extending lines represent the spread of the posterior distribution.

To examine the extent to which the observed effects are credible, I examine the posterior distributions of the estimates following best practices for hypothesis testing in the Bayesian framework: the Region of Practical Equivalence (ROPE) + Highest Density Interval (HDI) decision rule and Probability of Direction (PD; Kruschke 2014, 2018), described in detail below. These two measures are standard measures to evaluate the existence (Probability of Direction), and "significance" of an effect (ROPE + HDI). Significance here is a statement of the magnitude and certainty of the effect as reflected in the observed data and posterior estimates, and aids in the decision to accept or reject the effect but is not meant to be stand in for the frequentist evaluation of 'significance' testing. Given the effects observed above, both measures should provide additional clarity on the extent to which the effects in a given condition are credible and will allow us to compare the effects across the two conditions to determine the degree to which they appear to be the same underlying effects and pattern. In the next section, I will explain the HDI + ROPE in more detail, and then describe the findings of the model in reference to the decision rule. Following that, I will describe the Probability of Direction in more detail, and similarly describe the findings of the model in reference to the metric. I will then examine the marginal means of the interaction term to shed additional light on the effect, and then summarize.

The HDI + ROPE provides a criterion for deciding whether to accept, reject, or remain uncertain, for a particular effect as a function of the magnitude and credibility of the parameter values in relation to a null effect (Kruschke, 2014). The ROPE is an analyst specified range of values under which a parameter value would be equivalent to a null effect. The HDI summarizes the points of a distribution where the parameter coefficients have a higher probability density than points outside of the interval (i.e., a range of credible values), and is analogous to confidence intervals in null hypothesis significance testing. Rejecting a parameter value would equate to a null effect, whereas acceptance would equate to support for the effect of interest (i.e., the parameter coefficient). The HDI + ROPE evaluates what proportion of the HDI (i.e., the most likely coefficient values) falls within the ROPE (i.e., a null effect). For example, if the HDI is set to 89% (i.e., 89% of the area under the probability density curve) and 100% of that HDI falls within the ROPE, the parameter effect would be rejected (i.e., "non-significant"). The lower the percentage of the HDI that falls within the ROPE, the more credible the effect. In the following

evaluations, the HDI is set to 89% and the ROPE is set to a range of 0.18 to -0.18, in line with a small effect size for logistic regression coefficients (Makowski et al., 2019). The HDI is set to 89%, following standard conventions, as opposed to 95% because large samples are required to achieve stability of posterior estimates near the edges of the distribution (Kruschke, 2014; McElreath, 2014, 2018).

Figure 6.14 depicts the ROPE (as indicated with the blue highlight) and the posterior estimates, with the 89% HDI filled in grey (specific values are reported in Table 6.4). Several observations can be made from Figure 6.14 to evaluate the credibility of the effects. First, we see a slight positive interaction between Step and Test, whereby the odds of listeners /eɪ/ responses from pre-test to post-test increases as the continuum becomes more /ʊ/-like (β = 0.09). However, 87% of the HDI falls inside the ROPE (89% HDI = -0.07 – 0.24), indicating a null effect for this interaction (i.e., non-significance). This finding demonstrates that listeners are not making a targeted shift near the category boundary from pre-test to post-test. On the other hand, the negative effect of Test (β = -0.33) shows the odds of listeners /eɪ/ responses decrease from pre-test to post-test. Only 1% of the posterior estimates falling within the ROPE (89% HDI = -0.49 – -0.17), showing a credible effect of listeners' overall biasing responses away from /eɪ/ at post-test. Finally, we see a credible negative effect of Step (β = -1.75), whereby listeners /eɪ/ responses decrease as the continua become more /ʊ/-like, with 0% of the HDI falling within the ROPE (-2.48 – -0.46). Together these results provide evidence that listeners did not learn the exposure pattern, counter to the initial hypothesis. Rather, listeners appear to demonstrate a global decrease in the categorization of /eɪ/ tokens from pre-test to post-test and no evidence of a shift in their categorization boundary.

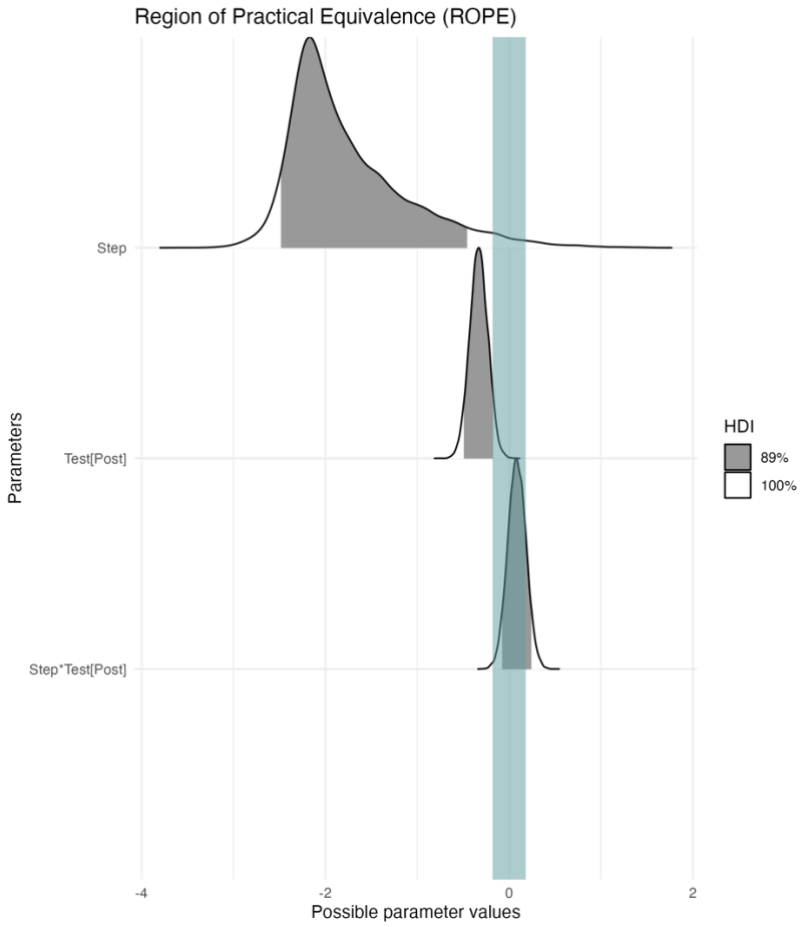

Figure 6.14 Estimated density distributions of the posterior shaded by High Density Interval (HDI). The blue shaded region represents the Region of practical equivalence (ROPE) with a range of -0.18 – 0.18.

Table 6.4 Parameter estimates and the proportion of the 89% HDI that falls within the ROPE (-0.18 – 0.18)

Parameter	**89% HDI**	**HDI + Rope %**
Intercept	[-1.13, -0.71]	0%
Step	[-2.48, -0.46]	0%
$Test_{[Post\text{-}Test]}$	[-0.49, -0.17]	1%
$Step*Test_{[Post\text{-}Test]}$	[-0.07, 0.24]	87%

As further clarification of the effects, Figure 6.15 depicts the Probability of Direction (PD), with the values reported in Table 6.5. Probability of direction, or the Maximum Probability of Effect, represents the certainty that an effect is observed in a particular direction—that is how much of the posterior distribution of the parameter coefficients are positive or negative (Makowski et al., 2019)[7]. In general, this can be interpreted as an index of effect existence but does not quantify evidence for or against the null hypothesis. As such, the probability of direction complements the HDI + ROPE decision rule which quantifies the magnitude and credibility of the effect in relation to the null (i.e., 'significance' of the effect). Figure 6.15 below shows the same distributions as above with the probability of positive and negative effects indicated by color or the proportion of the posterior distribution that shares the same sign as the median coefficient value. If more of the distribution is a single color, there is a higher probability that the observed effect is in a given direction and there are fewer observations of plausible effects in the opposite direction. For example, if we expect Step to have a negative effect on /eɪ/ responses (i.e., as step increases, /eɪ/ responses decrease), then the probability of direction will tell us what proportion of the sampled estimates conform to that expectation. If the probability is 100%, then all plausible coefficient values showed a negative effect. If the probability is 50%, then positive and negative effects are equally plausible, and we can't be certain that the observed effect (i.e., the reported median of the posterior) is the true effect.

Looking at Figure 6.15, overall, the effects align with the observations made above. Specifically, for the Step*Test term the majority of the posterior distribution is positive (PD = 80%) but the distribution crosses 0 and negative values are also likely. This suggests uncertainty in the effect estimate and provides further evidence that the interaction term is not credible given there is a 20% probability of a negative effect. To put it differently, the model estimates 80% of the time that listeners shift their category boundary towards the /ʊ/ end of the continuum, and the other 20% of the time estimates that listeners shift their boundary towards the /eɪ/ end of the continua. The probability for the Test term, however, is 99%, meaning almost all of the posterior estimates

[7] While the Probability of Direction may be interpreted in relation to the frequentist p-value, here I only use the metric as a means of capturing the uncertainty in estimating the effects as further evidence of the null or alternative hypothesis.

are negative, demonstrating greater confidence for the effect that the odds of listeners /eɪ/ responses decrease at Post-test. Finally, for the Step term, the probability of direction is 98%, with the majority of estimates being negative, indicating listeners decrease /eɪ/ responses as the continua become more /ʊ/-like. Overall, the probability of direction aligns with the HDI + ROPE in the previous section, showing that there is a credible effect of test, however not in the direction that constitutes learning. That is, the effect is the opposite behavior expected from the shift listeners were exposed to. However, based on the null effect of the interaction term (Step*Test), listeners are not showing a targeted shift of their categorization boundary towards the /eɪ/ end or the /ʊ/ end of the continua.

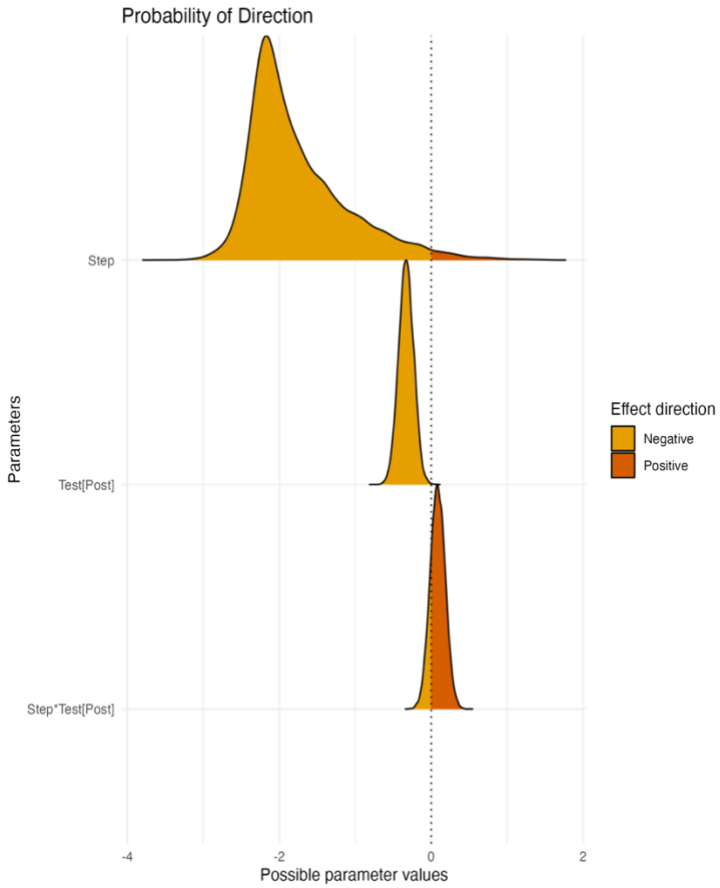

Figure 6.15 Estimated density function of the posterior estimates and color indicating the effect direction as indicated by the probability of effect.

Table 6.5 Probability of direction values for the posterior estimates, reflected in Figure 6.15 above.

Parameter	Probability of Direction	Direction
Intercept	100%	Negative
Step	98%	Negative
$Test_{[Post\text{-}Test]}$	99%	Negative
$Step*Test_{[Post\text{-}Test]}$	80%	Positive

Given the uncertainty in the interaction between Step and Test, and the predicted learning effect, I turn to briefly examine the marginal means for further clarity with the emmeans package (Lenth et al., 2023) in R (R Core Team, 2021). Figure 6.16 below presents the marginal posterior estimates of the interaction between Step and Test, with Step taken at the midpoint of the continuum (Step 0, from the scaled predictor). The marginal means below reinforce the patterns in Figure 6.14 and Figure 6.15, where at Step 0 the posterior estimates are slightly lower from pre-test to post-test. Indicating a boundary shift, at least near the center of the continua, from pre-test to post-test for the /eɪ/ condition, but with greater overlap between the posterior distributions. Altogether, the results in this section suggest that there is greater uncertainty for the estimates of step as a function of test, but a small credible effect for the effect of test. Meaning, the odds of listeners responding with more /eɪ/ tokens as the continuum becomes more /ʊ/-like (or vice versa) does not change from pre-test to post-test, but rather listeners show an overall decrease in /eɪ/ responses after exposure.

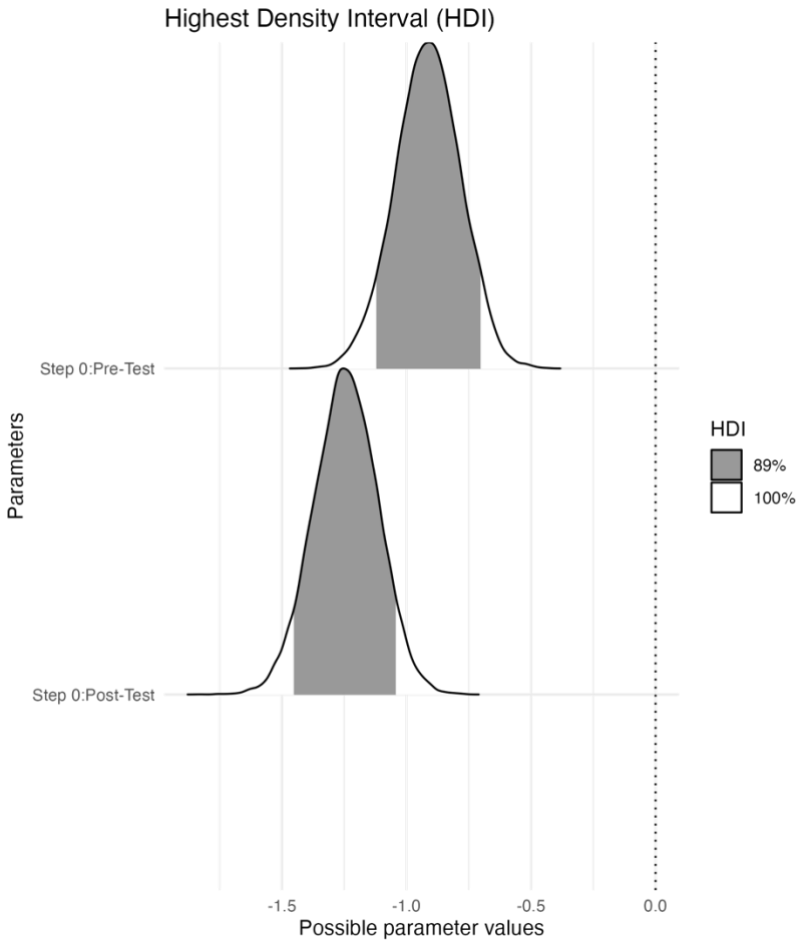

Figure 6.16 Marginal means 89% HDI based on Step 0 and level of Test. The mean of the distribution for Step 0: Pre-Test = -0.91, and Step 0: Post-Test = -1.24.

To summarize, the main effect of Step is a credible effect whereby the odds of listeners reporting /eɪ/ decreases as the continuum becomes more /ʊ/-like. The main effect of Test is credible and there is a decrease in /eɪ/ responses at post-test, meaning there is credible evidence that listeners did not learn the pattern of exposure but did demonstrate a change in categorization behavior at post-test. The slightly positive interaction between Step and Test does not appear to be a credible result with model estimates demonstrating greater uncertainty. Overall, these results suggest that listeners generally biased their responses away from /eɪ/ (i.e., towards /ʊ/) during post-test categorization and did not demonstrate a targeted shift in their category boundaries. Such a finding may point towards greater uncertainty for listeners at post-test; I will return to this point in the broader discussion and consider some possible explanations for this effect.

8.2.2.2 /ʊ/-Biased Condition:

Moving on to next analysis, I examine the main effects of the /ʊ/-Biased condition. The modeling choices outlined in the previous section (/eɪ/-Biased Condition) were repeated here, but on the subset of data for the /ʊ/-Biased condition and with the positive response variable representing /ʊ/ responses and Step reversed (as described above). In this section I will again first describe the overall effects and then follow the same analytic choices for hypothesis testing as the previous sections (i.e., HDI + ROPE and PD). First, as depicted in Figure 6.17, the /ʊ/ model here validates the full model in Section 8.2.1 above, demonstrating an overall increase in /ʊ/ responses from Pre-Test to Post-Test (β = 0.76, *SE*= 0.11, 95% CI = 0.54 – 0.99). In addition, there is a small effect for the interaction of Step and Test (β = 0.16, SE= 0.09, 95% CI = 0.0 – 0.35), indicating an increase in /ʊ/ responses at post-test as step increases, or in other words a positive shift in the boundary at post-test where listeners report more /ʊ/ responses as the continuum becomes more /eɪ/-like. Finally, the main effect of step illustrates a negative effect, demonstrating that as step increases odds of /ʊ/ responses decrease (β = -1.89, *SE*= 0.70, 95% CI = -2.77 – -0.15).

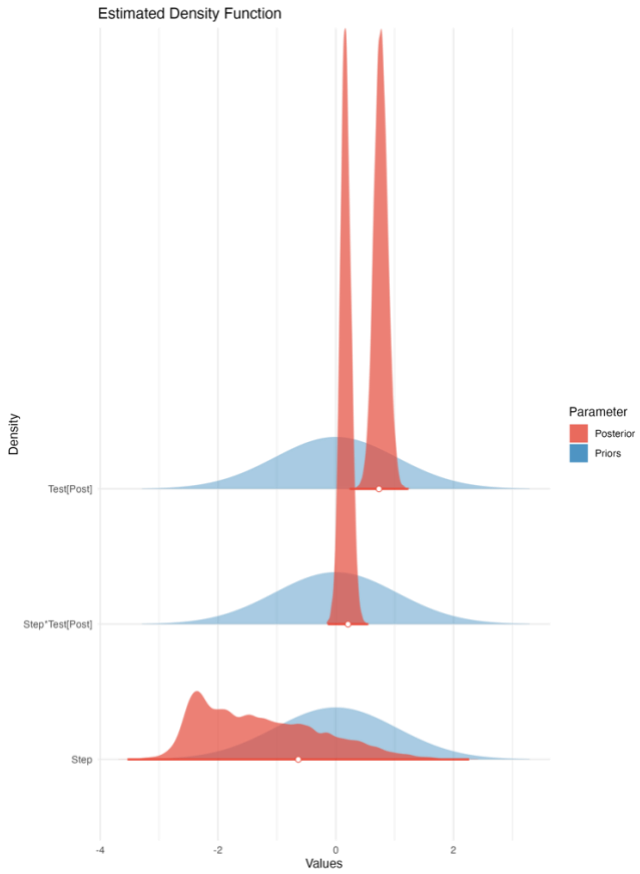

Figure 6.17 Estimated density function of posterior estimates (red) overlaid the estimated density function of the prior distribution (blue).

Figure 6.18 depicts the posterior distributions of the fixed effects in the model, using the HDI + ROPE decision rule. As above, the criteria are an 89% HDI and a ROPE indicating a small effect, from -0.18 – 0.18. Figure 6.18 and Table 6.6 illustrates a credible effect of Test, with 100% of the HDI falling outside of the ROPE (0.54 – 0.99). These results demonstrate a reliable effect, providing evidence of learning in the direction of exposure. In contrast, the interaction of Test and Step demonstrates an effect with less credibility, with 60% of the HDI falling within the ROPE (-0.01 –0.34). This effect demonstrates that there is not a credible effect of listeners shifting their boundary at post-test, or the effect is minimally different from a null effect. Additionally, the /ʊ/ model illustrates similar properties as the /eɪ/ model in terms of the effect of Step, with a high degree of variability in model estimates despite 0% of the HDI falling within the ROPE (-2.83 – -0.29). There is a credible effect of Step, with greater magnitude than

other effects in the model, showing that listeners decrease their /ʊ/ responses as the continuum becomes more /eɪ/ like (i.e., Step increases). Thus far, these results are in line with the results of the /eɪ/-Biased condition, showing that each condition patterns similarly, with credible effects of Test and Step, but only weak evidence of an interaction between the two. In the case of the /ʊ/-Biased condition, however, listener behavior follows from the pattern of exposure, and is of greater magnitude, suggesting learning occurs for the /ʊ/-Biased condition.

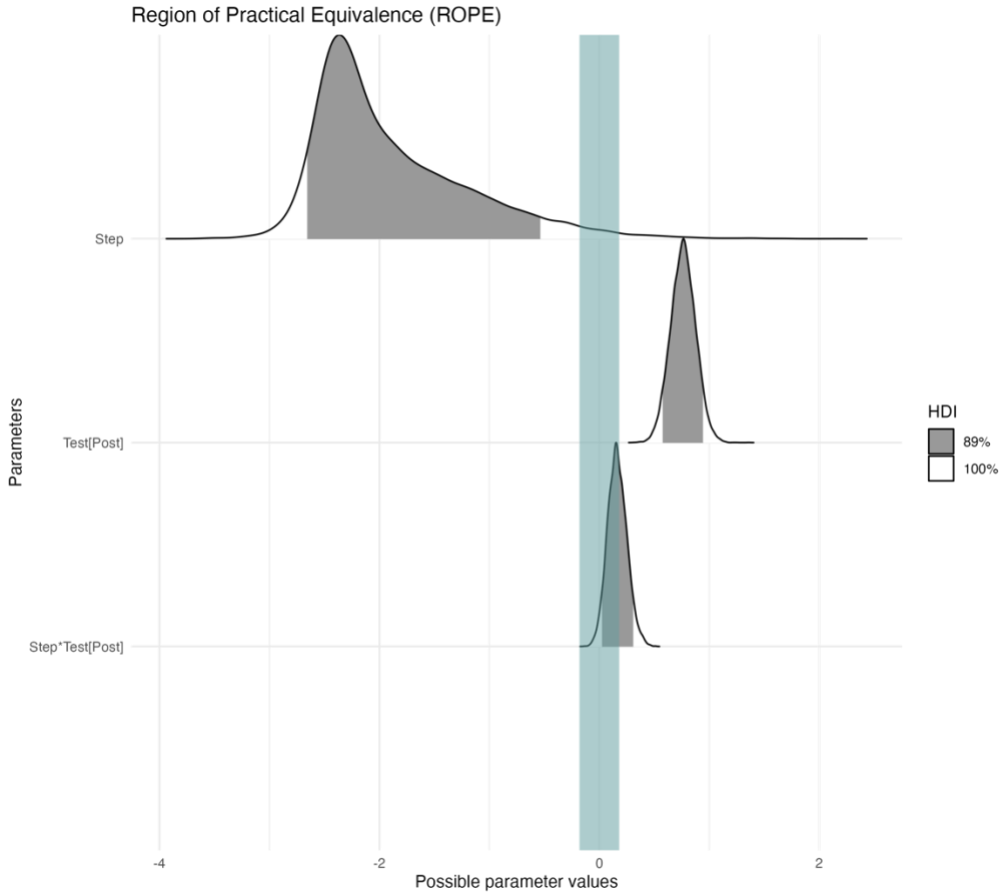

Figure 6.18 Estimated density distributions of the posterior shaded by High Density Interval (HDI). Blue shaded region represents the Region of practical equivalence (ROPE) with a range of -0.18 – 0.18.

Table 6.6 Parameter estimates and the proportion of the HDI that falls within the ROPE.

Parameter	89% HDI	HDI + Rope %
Intercept	[0.66, 1.27]	0%
Step	[-2.83, -0.29]	0%
$Test_{[Post\text{-}Test]}$	[0.54, 0.99]	0%
$Step*Test_{[Post\text{-}Test]}$	[-0.01, 0.34]	60%

To further clarify the effects, we again look at the Probability of Direction. As illustrated in Figure 6.19 and Table 6.7, the posterior distributions illustrate 100% probability of a positive effect for the main effect of Test, and a 97% probability of a positive effect for the interaction between Test and Step, representing a moderate positive effect. Such a finding suggests that despite the effect appearing small and falling within a null range, there is high confidence in the direction of the effect. The effect here contrasts with the finding for /eɪ/ where the same interaction shows higher uncertainty in the effect of the term. Overall, this may suggest that listeners are more likely shifting their boundary towards the /eɪ/ end of the continuum in the /ʊ/-Biased condition, as expected with learning the exposure shift. Finally, the model demonstrates a high degree of certainty around the main effect of step with 98% probability of a negative effect. Based on the posterior estimate distributions and the relative strengths outlined here, I argue that the effect of an increase at post-test in /ʊ/ responses is credible. Similarly, the effect of Step demonstrates credible effect given the magnitude of the effect and the absence of overlap with the ROPE and the high probability of direction. To further elucidate the effect of the Step and Test interaction, I once again look at the marginal means using the emmeans package (Lenth et al., 2023) in R (R Core Team, 2021).

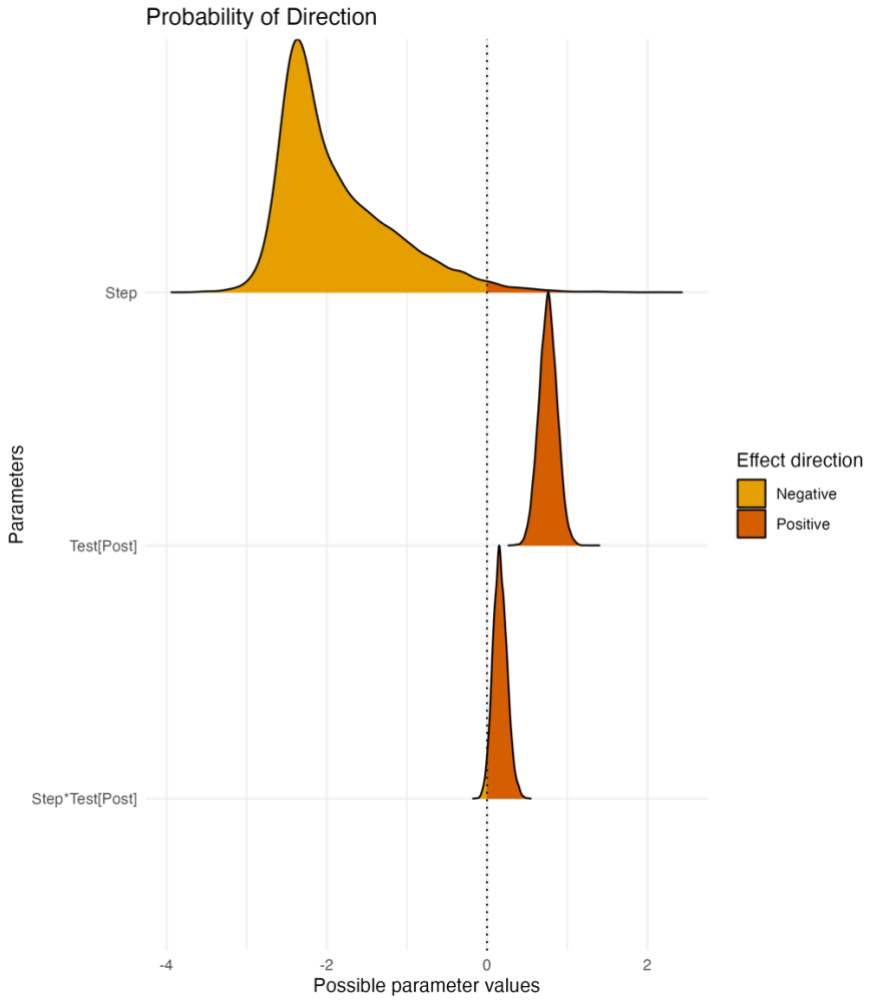

Figure 6.19 Estimated density function of the posterior estimates and color indicating the effect direction as indicated by the probability of effect.

Table 6.7 Probability of direction of the effects, as illustrated in Figure 6.19.

Parameter	**Probability of Direction**	**Direction**
Intercept	100%	Positive
Step	98%	Negative
$Test_{[Post-Test]}$	100%	Positive
$Step*Test_{[Post-Test]}$	97%	Positive

Looking at the marginal means in Figure 6.20 we can see that there is indeed an increase in the log-odds of /eɪ/ responses from pre-test to post-test in the middle of the continua (Step 0, again the middle of the centered step predictor). The 89% HDI shows no overlap in the pre-test posterior estimates compared to the post-test posterior estimates, and there is little variability. While Step 0 shows an increased likelihood of /ʊ/ responses during pre-test, we see the effect strengthen when we get to the post-test. While the overall effect appears to be small, and indeed requires further research, the probability of direction coupled with the marginal means here suggest the effect is qualitatively different for the /ʊ/-Biased condition compared to the /eɪ/-Biased condition. The marginal means more clearly supports the fact that listeners have shifted their boundaries in response to the exposure shift in the /ʊ/-Biased condition, as the magnitude of the posterior estimates and their lower degree of overlap contrast the /eɪ/-Biased condition.

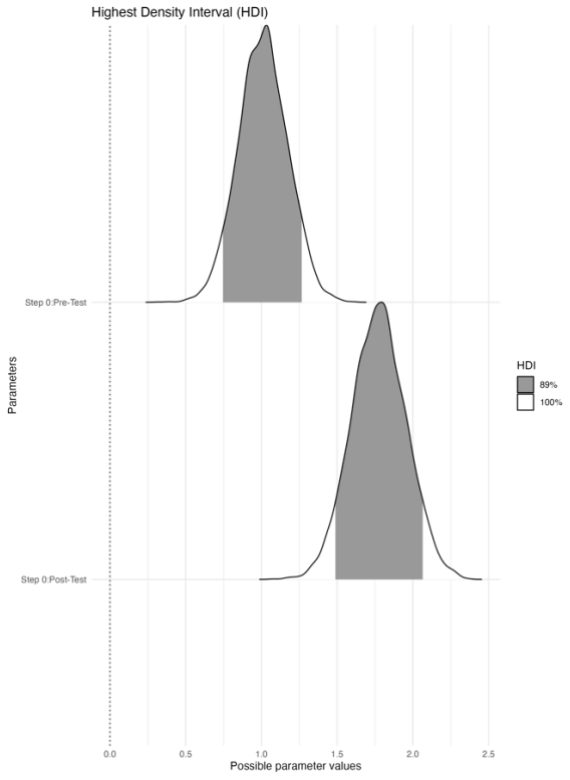

Figure 6.20 Marginal means 89% HDI based on Step 0 and level of Test. The mean of the distribution for Step 0: Pre-Test = 1.01, and Step 0: Post-Test = 1.78.

8.2.3 Interim Discussion

The learning results of the experiment do not align with the original hypotheses. For review, the two vowel categories were chosen for this experiment based on the fact that they were critically different concerning what level of socio-indexical information was informative of their cue distributions. I hypothesized that both conditions would demonstrate learning, whereby /eɪ/ responses would increase at post-test for the /eɪ/-Biased condition, and /ʊ/ responses would increase at post-test for the /ʊ/-Biased condition. While the /ʊ/ condition conformed to these expectations, the /eɪ/-Biased condition showed a decrease in /eɪ/ responses at post-test. While these results show an /ʊ/ preference across both conditions, the results above suggest the effects are qualitatively different. In particular, the /ʊ/-Biased condition demonstrates higher confidence in the estimates of Test and the interaction of Test and Step in contrast to the same predictors observed in the /eɪ/-Biased condition. Higher confidence in the predictors suggests that listeners more robustly shifted their boundaries in the /ʊ/-Biased condition compared to the /eɪ/-Biased condition. Similarly, the marginal effects of the /eɪ/-Biased condition show no qualitative difference between the posterior distribution, showing no evident shift, at least in the middle of the continua. On the other hand, the marginal means of /ʊ/ show a very clear increase in the log odds from pre-test to post-test at the same point in the continua. Overall, the effects in the /eɪ/-condition I interpret as listeners increasing in *uncertainty* from pre-test to post-test. On the other hand, the effects of the /ʊ/ condition provide evidence of perceptual learning of the exposure shift. I will return to this interpretation for further refinement in the larger discussion section (Section 9) following the generalization results.

8.3 Generalization

I now turn to the second question of this experiment: is there an asymmetry in generalization across the two vowel categories? The two novel generalization talkers represent a same gender pair with the exposure talker (a female talker T2_F) and a different gender pair (a male talker T3_M). I hypothesized that the dialect-informative vowel category (/eɪ/-Biased condition) would exhibit robust cross talker generalization, with the pattern extending to both the novel male and female talkers. On the other hand, I hypothesized there would be no

generalization for the talker-informative condition (/ʊ/-Biased condition) or it would be limited to the same gender pairing as a function of acoustic similarity.

Given the unexpected learning results in Section 8.2 above, it is unclear whether we would see the robust generalization predicted for the /eɪ/ condition. The original hypothesis hinges on listeners *learning* the exposure pattern and generalizing the learned pattern. However, since listener behavior changed, but was not evidence of learning the pattern from exposure, it would seem reasonable that generalization of the shift in their behavior from pre- to post-test would be limited or non-existent. On the other hand, it's possible that listeners may generalize their updated beliefs about the category regardless of whether they demonstrated *learning* of the novel pattern. More precisely, if listeners in the /eɪ/-biased condition became more uncertain about the category, and not just the talker-specific percepts, users may extend their uncertainty to similar talkers. Given that they were as expected, the learning results for /ʊ/ do not provide any explicit need to revisit the hypotheses. However, given the results of /eɪ/, we might posit that the link to socio-indexical factors may not correlate with listener behavior as predicted, and as such this could extend to different generalization behaviors than hypothesized in the /ʊ/-Bias condition by, for example, generalizing to other talkers. I will return to some of these issues more in the discussion following the results. I will examine the raw results first and move forward to a statistical analysis for support and further refinement of the observed trends.

As a reminder, generalization talkers were counterbalanced across participants within each condition, so that individual participants were only tested on either the female talker (T2_F) *or* the male talker (T3_M) following post-test of the exposure talker. Additionally, learning was evaluated by evidence of change in listener behavior in the direction of exposure. However, generalization may be demonstrated by extending the behavior from the exposure talker at post-test, as in a lower response of /eɪ/, across both conditions. In the analyses below, the Step predictor is coded so that each condition follows the same response variable, /eɪ/ responses, with Step 1 indicating the most /eɪ/-like end of the continua and Step 7 indicating the most /ʊ/-like end. Any positive slope for Step suggests an increase in /eɪ/ responses as the continua become more /ʊ/-like, and a negative slope of Step suggests a decrease in /eɪ/ responses as the continua become more /ʊ/-like.

Figure 6.21 shows the raw results of generalization across the two novel talkers along with the exposure talker for reference. In Figure 6.21, we see that there is an overall difference between the two conditions for the male talker, such that we see overall fewer /eɪ/ responses in the /ʊ/-Biased condition compared to the /eɪ/-Biased condition. However, there is no difference between the conditions for the female talker. These results suggest generalization to the male talker for the /ʊ/-Biased condition, but no generalization to the female talker in either condition. Below I will assess the credibility of the experimental effects through statistical analysis followed by a brief discussion of the relationship between the talkers and their baseline categorization function from the norming phase of the study for further elucidation of these effects.

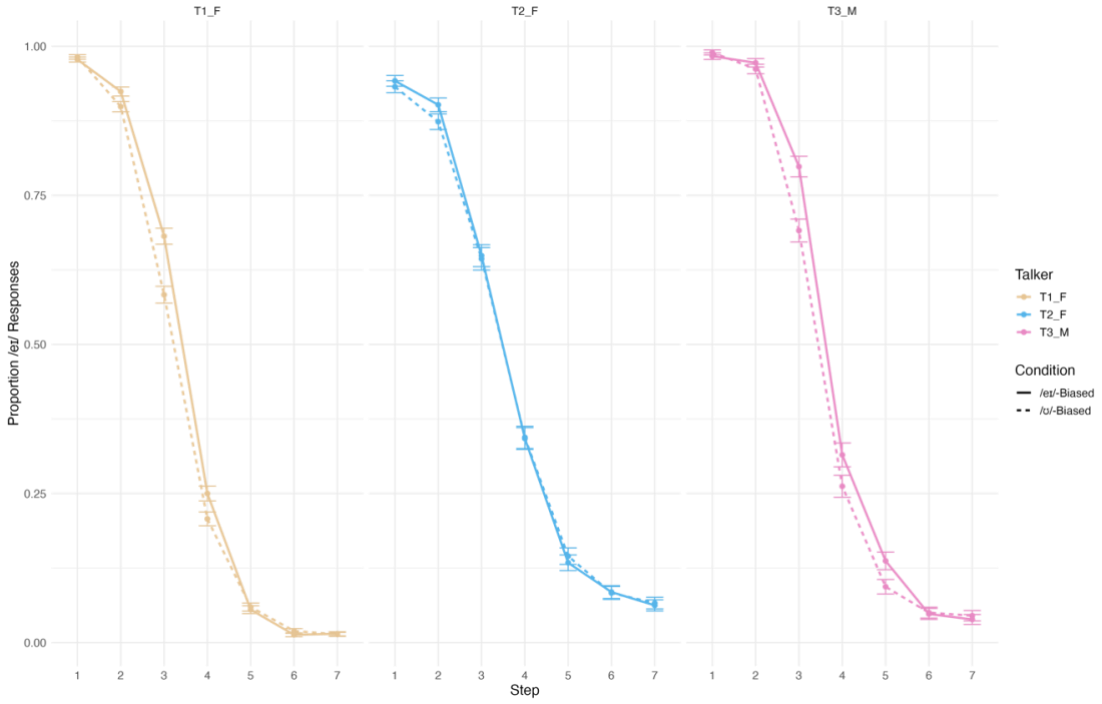

Figure 6.21 Proportion of /eɪ/ responses for each Step of the continua. Step is depicted as factor for interpretation purposes. Line colors represent talker (Generalization talkers = T2_F and T3_M; Exposure talker = T1_F) and line type represents condition. Error bars indicate standard error.

To test whether generalization differed as a function of condition (/eɪ/-Bias or /ʊ/-Biased) and generalization talker (male T3_M or female T2_F), this section examines only the post-test categorization data across the two generalization speakers. In line with the learning section above

(Section 8.2), I ran Bayesian logistic regression in the brms package (Bürkner, 2017) in R using the Bernoulli distribution and a logit link function, with 4 chains and 8000 iterations per chain, with 4000 warm-up iterations (half of the total iterations per chain). The dependent variable is a binary coded response with 1 indicating an /eɪ/ response from participants, and 0 indicating an /ʊ/ response. The independent variables were talker, condition, and step, and their interactions. Speaker and Exposure were treatment coded; talker T2_F was the reference level for the talker effect and the /eɪ/-Bias condition as the reference level for condition. Step was entered as a scaled and centered numeric variable. All main effects and their interactions were specified with weakly informative prior (*Normal*, $\mu = 0, \sigma = 1$). The random effect structure included random slopes for all main effect interactions over participants and a random slope for step by item. All random effect priors were specified with a regularizing prior (*Normal*, $\mu = 0$, $\sigma = 1$), and correlations terms were fit with an LKJ prior. Models were inspected visually and through model diagnostics to evaluate convergence; chains have converged ($\hat{R} = 1.0$) and there is no evidence of high collinearity between predictors.

Following the previous sections, I examine the posterior distributions of the estimates in reference to the HDI + ROPE decision rule and the probability of direction (Kruschke, 2014, 2018), provided in

Table 6.8 and Table 6.9. Figure 6.22 depicts the model's posterior estimates and the 89% High Density Interval (HDI) and the ROPE, and Figure 6.23 depicts the Probability of Direction (PD) of the posterior estimates. As in the overall model in Section 8.2, there is a credible bias towards /ʊ/ responses (β = -0.64, *SE*= 0.16, 95% CI = -0.95 – -0.35), with 0% HDI + Rope and 100% negative PD. Similarly, there is a large effect of Step, where the odds of /eɪ/ responses decrease as step increases becoming more /ʊ/-like (β = -2.45, *SE*= 1.20, 95% CI = -4.59 – 0.46), with 0% HDI + Rope and 96% . There is a small effect of condition, whereby the odds of /eɪ/ responses are lower in the /ʊ/-Biased condition compared to the /eɪ/-Biased condition (β = -0.05, *SE*= .22, 95% CI = -0.48 – 0.39), indicating fewer /eɪ/ responses overall in the /ʊ/-Biased condition (0% HDI + ROPE and 100% negative PD). Finally, there is a slightly positive effect of talker, where there is an increase in the odds of /eɪ/ response for the male talker (T3_M) compared to the female talker (T2_F; β = 0.12, *SE*= 0.23, 95% CI = -0.33 – 0.59). In other words, listeners were more biased to respond with /ʊ/ for the novel female talkers' continua and for increasingly /eɪ/-like steps along the continua compared to the male talker.

However, the three-way interaction of Step, Condition, and Talker demonstrate a higher degree of uncertainty in the effect with 36% of the posterior HDI falling within the ROPE and only 74% negative PD (β = -0.24, *SE*= 0.37, 95% CI = -0.98 – 0.46), suggesting a non-credible effect and interpretation of the null. In other words, there is minimal difference between the listeners' generalization behavior of either talker across the two conditions. This finding illustrates that listeners in both conditions demonstrated the same pattern of generalization to both the same-gender and different-gender paired talkers and did not show shifts in their categorization boundary. There is no support for the original hypothesis that the two conditions generalization patterns would differ. However, given the multiple levels for the interaction term, there may be a difference between certain comparison levels (e.g., only one talker, one condition, compared to the others). To better understand the three-way interaction, I turn to examine the marginal means below to better determine whether each talker and condition are comparable at different points along the continua. Namely, this seeks to identify whether the effect of the three-way interaction is obscuring the effect for a single talker and condition or whether the talkers are comparable across conditions along the continua.

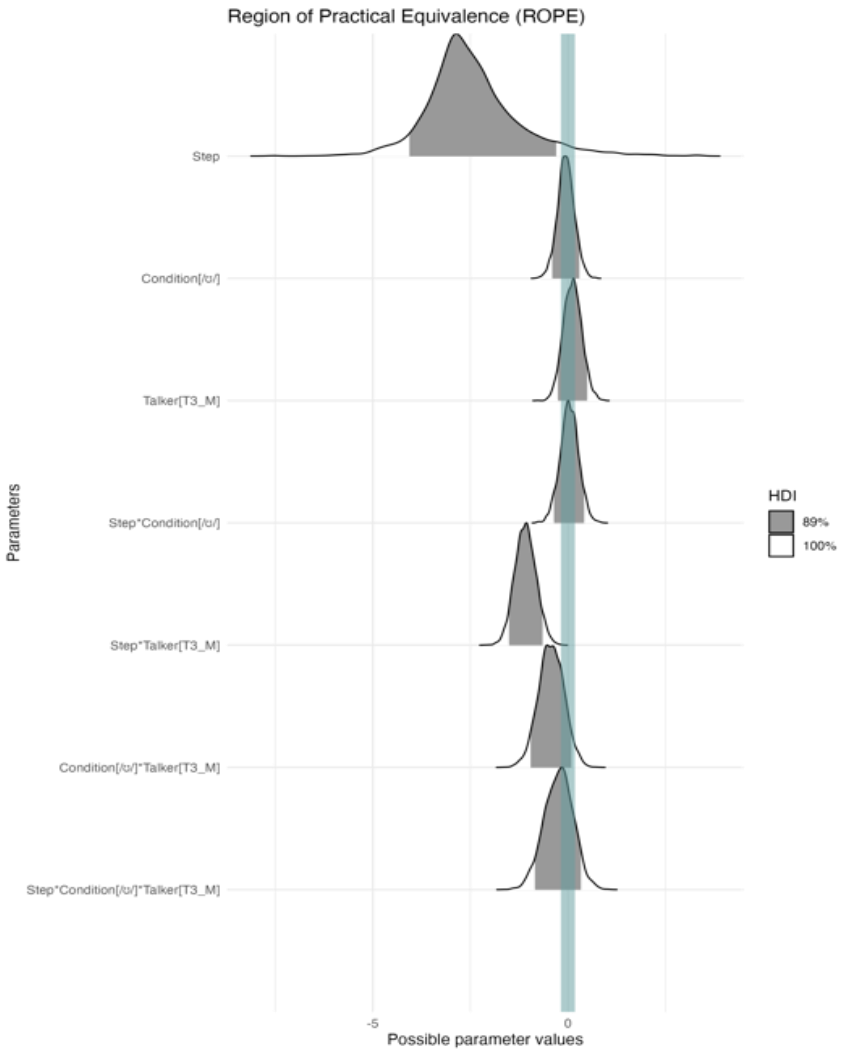

Figure 6.22 The HDI + ROPE decision rule. Shaded region representing the 89% HDI and the blue shaded region representing the ROPE at -0.18 – 0.18.

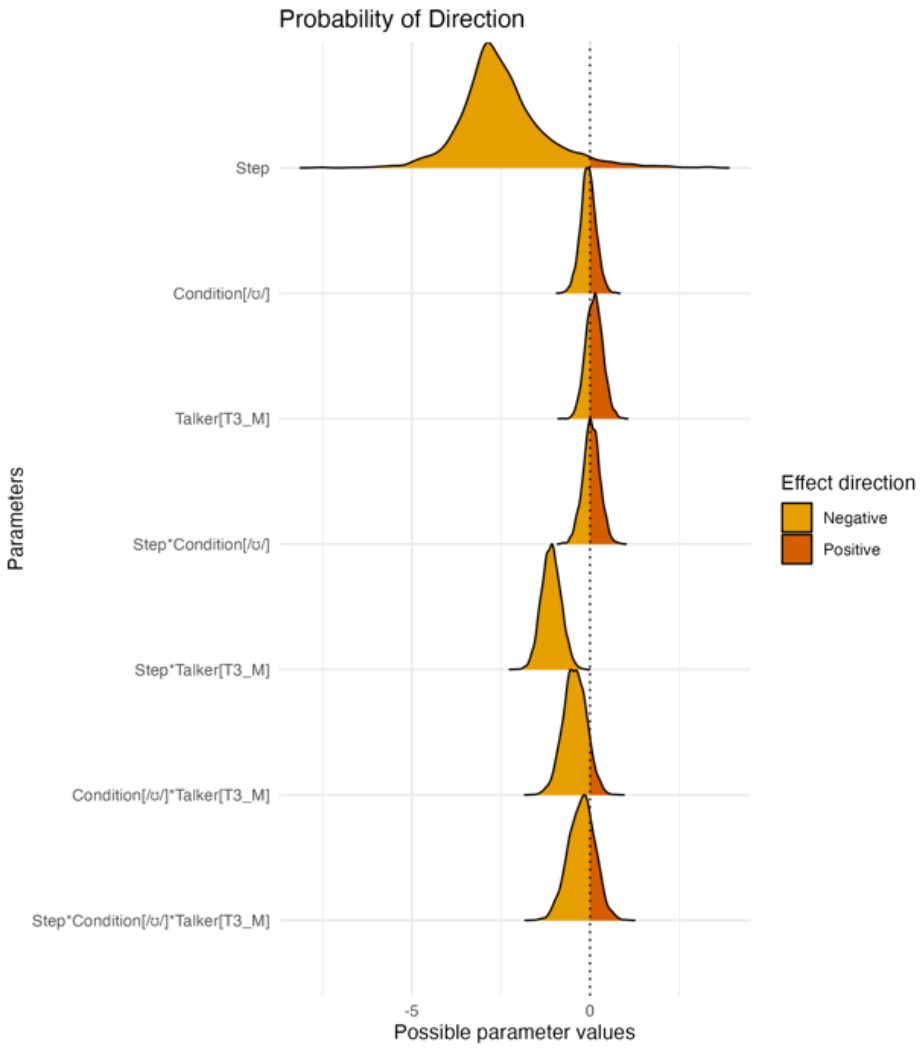

Figure 6.23 Probability of direction, with color indicating direction of effect.

Table 6.8 Parameter and percentage of ROPE within the HDI for effects illustrated in Figure 6.23

Parameter	89% HDI	ROPE Percentage
Intercept	[-0.88, -0.39]	0%
Step	[-4.22, -0.53]	0%
$Condition_{[uh\text{-}Bias]}$	[-0.38, 0.31]	67%
$Talker_{[T3\_M]}$	[-0.24, 0.51]	57%
$Step*Condition_{[uh\text{-}bias]}$	[-0.33, 0.44]	63%
$Step*Talker_{[T3\_M]}$	[-1.51, -0.65]	0%
$Condition_{[uh\text{-}Bias]}*Talker_{[T3\_M]}$	[-0.93, 0.11]	20%
$Step*Condition*_{[uh\text{-}Bias]}*Talker_{[T3\_M]}$	[-0.80, 0.37]	36%

Table 6.9 The probability of direction for effects illustrated in Figure 6.23

Parameter	Probability of Direction	Direction
Intercept	100%	Negative
Step	96%	Negative
$Condition_{[uh\text{-}Bias]}$	60%	Negative
$Talker_{[T3\_M]}$	69%	Positive
$Step*Condition_{[uh\text{-}bias]}$	57%	Positive
$Step*Talker_{[T3\_M]}$	100%	Negative
$Condition_{[uh\text{-}Bias]}*Talker_{[T3\_M]}$	91%	Negative
$Step*Condition*_{[uh\text{-}Bias]}*Talker_{[T3\_M]}$	74%	Negative

Given the interest in the individual talkers across each condition, I move forward to look at the marginal effects using the emmeans package (Lenth et al., 2023) in R (R Core Team, 2021). Figure 6.23 plots the conditional marginal posterior estimates for the three-way interaction of Step, Condition, and Talker, taking the center of the Step predictor for estimates.

This figure elucidates the patterns above by demonstrating that T3_M and T2_F appear to have similar posterior estimates for the /eɪ/ condition, and the estimates for T2_F show no difference between the two conditions. The only estimate that suggests a credible difference is the center of the continuum (Step 0) for the /ʊ/ Bias condition and T3_M. In other words, there is a decrease in the log-odds of /eɪ/ responses for the male talker only in the /ʊ/-Biased condition, illustrating that listeners appear to have generalized the exposure pattern in the /ʊ/-Biased condition to the male talker, but not the female talker. Similarly, listeners do not appear to have generalized the change in behavior from the /eɪ/-Biased exposure to either the male or female talkers.

The generalization results are surprising and contrary to the hypotheses described in Section 2 (Predictions). However, given the direction of the effect for learning in the /eɪ/-Biased condition (i.e., reduced /eɪ/ responses), it is unclear whether the behavioral pattern of the exposure talker at post-test was extended to the novel talkers. From the results thus far we cannot confirm whether listeners generalized the same reduction in /eɪ/ responses from the exposure talker in the /eɪ/-Biased condition or reset their categorization behavior for the new talkers. If listeners generalized their beliefs, demonstrating greater uncertainty of the category more globally, it may explain what appears to be an absence of generalization to the female talker, as listeners may have fewer /eɪ/ responses for the novel female talker in both conditions. Correspondingly, the categorization of the male talker may be reflective of generalization for the /ʊ/-Biased condition but talker-specific learning for the /eɪ/-Biased condition, or greater magnitude of shift for the /ʊ/-Biased condition compared to the /eɪ/-Biased condition.

In other words, it's unclear whether listeners demonstrate generalization of updated beliefs about the category or talker-specific changes to categorization behavior (i.e., only the exposure talker). In order to elucidate this concern, I return to the norming data below to compare responses across the experiment and norming data. As the current experiment did not elicit baseline responses to the generalization talkers, the norming study can facilitate interpretation by acting as a proxy 'baseline' representation of listeners' categorization functions. Listeners in the norming study were not provided any exposure shifts before participating in categorization and should thus shed light on the degree to which listeners behave differently across the exposure conditions. There is, however, the caveat that listeners in the norming

experiment were exposed to more steps along the continua, which may have influenced their categorization function, but the results should nonetheless help in disambiguation of the pattern.

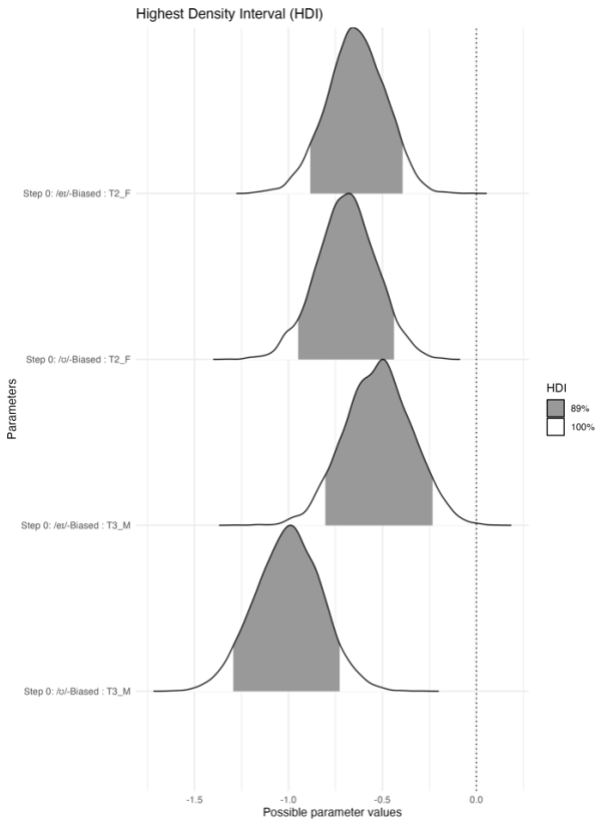

Figure 6.24 Posterior estimate distributions of the marginal coefficients. Shaded region represents 89% HDI.

Figure 6.25 below shows the raw categorization functions across both talkers across three conditions: Norming, /eɪ/-Biased, and /ʊ/-Biased. Here we can see that the two talkers demonstrate different patterns across the three experimental contexts. Based on visual inspection of the curves in Figure 6.25, the responses for the three conditions are similar for talker T2_F, showing that no generalization occurred with the novel female talker. Additionally, the response curve suggests there is more uncertainty in categorization for talker T2_F, as evidenced by the shallower categorization slope whereby even the most extreme ends of the continua lack categorical /eɪ/ or /ʊ/ responses. However, when examining the male talker, T3_M, generalization appears to have occurred in both conditions. In the /eɪ/-Biased condition, there are

fewer /eɪ/ responses compared to the Norming data, and the fewest /eɪ/ responses in the /ʊ/-Biased condition which parallels the patterns observed for the exposure talker (T1_F) across the conditions. The male talker also demonstrates a steeper slope in categorization compared to talker T2_F with more categorical responses at the end points. In other words, listeners appear to have generalized from the exposure talker to the novel male talker but not the female talker across both conditions.

These results are surprising and counter to the initial hypothesis that cross-talker generalization should occur for the /eɪ/-Biased condition but not the /ʊ/-Biased condition or restricted to same-gender pairs. The results provide counter evidence to previous work suggesting listeners should be more likely to generalize to talkers of the same gender for vocalic shifts (e.g., Kleinschmidt, 2019) as a result of gross acoustic differences across genders. A potential explanation for this trend is the perceptual similarity of the stimuli despite the acoustic patterns. Namely, it possible the exposure talker and the male talker's test continua are perceptually similar while the other female talkers' continua are dissimilar, despite the acoustic similarity of the two female talkers. The categorization results of the three talkers from norming appear to reflect this perceptual disparity, with the exposure talker (T1_F) exhibiting a similar slope and boundary point to the male talker (T3_M) but dissimilar to the other female talker (T2_F). The observations presented here may lend support to the proposal that cross-talker generalization is restricted to perceptually similar segmental ranges, even when the acoustic similarity of the talkers is similar (Reinisch & Holt, 2014). I will return to this point in the discussion in more detail and its relationship with the learning results above.

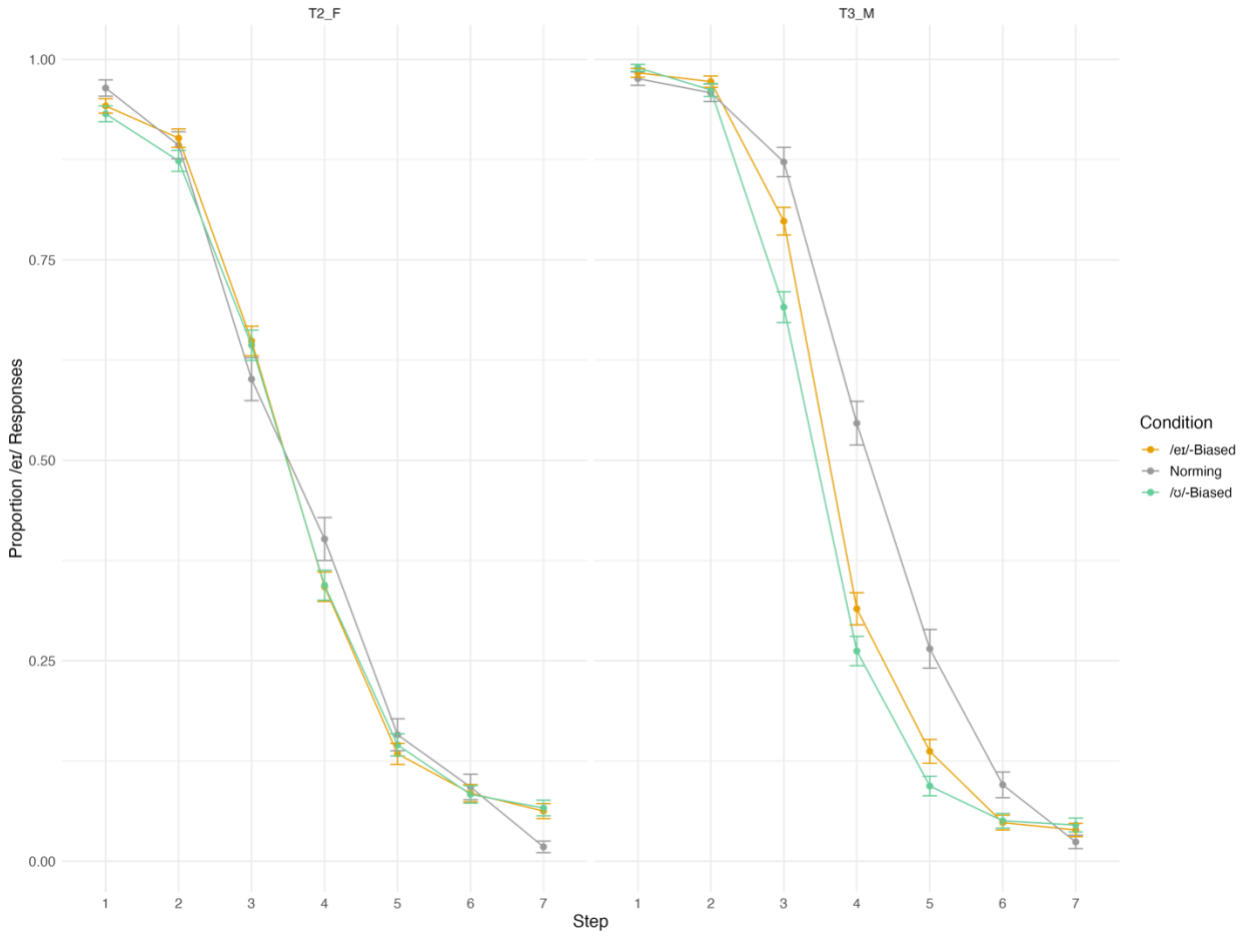

Figure 6.25 Proportion of /eɪ/ responses for each Step of the continua faceted by talker. Step is depicted as factor for interpretation purposes. Line colors represent condition. Error bars indicate standard error.

9 Discussion

Overall, the results of this experiment pose interesting challenges for ideal adapter models and perceptual learning and generalization more broadly. The above results illustrate asymmetrical learning, such that the /ʊ/ → /eɪ/ shift was learned, but the /eɪ/ → /ʊ/ shift was not, with listeners exhibiting an opposite pattern with decreased /eɪ/ responses at post-test. Listeners also generalized the learned /ʊ/ → /eɪ/ shift to the novel male talker, but not the novel female talker. Correspondingly, listeners generalized the reduction in /eɪ/ responses in the /eɪ/ → /ʊ/ shift for the male talker, but again not the female talker. The generalization results of the /ʊ/-Biased condition (/ʊ/ → /eɪ/) refute the original hypothesis that listeners would show less

generalization as a consequence of the category being linked to individual variation but not group variation (i.e., talker-informative). Relatedly, the dialectally informative category (/eɪ/) does not support the hypothesis that listeners would show greater remapping and generalization as a result of the category being linked to dialect variation (i.e., dialect-informative).

The results of the /ʊ/-Biased condition highlight that listeners learn the novel shift of /ʊ/ towards /eɪ/ and generalize the novel vocalic pattern to perceptually similar talkers. The fact that listeners learned the exposure shift aligned with the original predictions about learning for this category. However, the generalization results do not conform to the original predictions, as listeners appear to generalize to other talkers, and it is not restricted to the same-gender talker. This suggests that the *talker-informative* hypothesis does not adequately describe listener behavior for this vowel category. Rather, it appears that listeners update their beliefs about the category boundary following exposure and generalize it to other novel talkers with perceptually similar segmental ranges. Such a finding refutes claims that learning is talker-specific when the segment contains a high degree of talker-specific detail, such as spectral information (Kraljic et al., 2008; Kraljic & Samuel, 2006) or is idiosyncratically conditioned (Kleinschmidt, 2019; Kraljic et al., 2008). The results of the /eɪ/-Biased condition further suggest that the asymmetry in perceptual learning and generalization of vowels is unlikely to directly follow from the initial hypothesized talker-informative and dialectally informative dichotomy predicted. In the following section, I will discuss possible explanations for the results of the /eɪ/ condition in more detail focusing primarily on the learning results, then turn to the generalization results, and return to the intersection therein to conclude.

9.1 Learning

In this section I aim to specifically address the surprising result of the absence of learning and decrease in /eɪ/ responses in the /eɪ/-Bias condition. Overall, I will argue that listeners are demonstrating an increase in uncertainty following exposure to the novel shift in /eɪ/. Given the results, the original dichotomy between talker-informative and dialect-informative does not adequately explain listeners' perceptual learning behavior. While there are any number of potential explanations for this effect, I focus primarily on two plausible aspects of distributional properties that may drive the findings. First, I discuss the possibility that short-term distributional

characteristics of the stimuli caused the reduction in /eɪ/ responses during post-test. I argue that short-term characteristics are unlikely to be the cause of listener behavior. In the subsequent section, I consider how listeners' knowledge of long-term distributional characteristics of the vowel category and causal ambiguity provide a better account for the observed behavior. Given this discussion, the socio-indexical structure linked to each category and its role in perceptual learning needs to be revised, which I will return to at the end of the section.

9.1.1 Short-Term Distributional Properties

Short-term distributional properties provide a possible avenue for explaining the reduction in /eɪ/ responses at post-test. One may hypothesize that the decrease in /eɪ/ responses at post-test is driven by asymmetrical variance of the exposure items across the two conditions. This hypothesis is derived from previous work demonstrating that listeners' perceptual learning behavior is influenced by the degree of variability they are exposed to (Babel et al., 2019; Sumner, 2011). In addition, some work predicts that listeners exposed to wider distributions of a category will have greater uncertainty in categorization (Clayards et al., 2008; Theodore & Monto, 2019). If listeners in the /eɪ/-Biased condition were exposed to more variable stimuli than those in the /ʊ/-Biased condition, their post-test behavior may be indicative of this same mechanism. However, the estimated variance across F1 and F2 for both categories during exposure (depicted in Figure 6.8 in Section 6), shows approximately equal variance across the conditions for these cues, suggesting this explanation is unlikely the cause of listener behavior.

An alternative explanation is that the distributional properties led listeners to reweight cue expectations, shifting their reliance to other cues for categorization at post-test. Such an explanation is supported by previous research demonstrating listeners are able to learn distributional characteristics of individual cue dimensions and reweight cues according to novel talker patterns (Idemaru & Holt, 2011, 2014; Liu & Holt, 2015). For example, Liu and Holt (2015) demonstrate that in a categorization task listeners adjusted the relative weighting of acoustic cues to a vowel contrast based on the distributional properties of the stimulus set. A similar effect may account for listeners' behavior in the /eɪ/-Biased condition here, whereby listeners reweight attention to F1 as a cue over F2. Cue reweighting in this case may be driven by the fact that on average F1 is lower in the exposure stimuli than the categorization stimuli, and

variability is slightly larger in F2 than F1 in exposure. The categorization items mostly approach the F1 values of the stimulus set at the most extreme steps (1 & 2). As a result, it's possible that as listeners are exposed to items that *predominantly* vary in F2 (i.e., items are backing) they reweight expectations towards F1 because F2 has become an unreliable cue to /eɪ/. During categorization, the categorization stimuli do not represent extreme enough points of F1 for listeners to be confident that the speaker produced /eɪ/. It's also possible that rather than listeners reweighting cues, the distributional properties perhaps only magnified the reliance on F1, since the categorization items at pre-test have a strong bias towards /ʊ/ before exposure to the novel shift.

I argue that either explanation is unlikely given that the reduction in /eɪ/ responses was generalized to the male talker, whose categorization items maintain relatively stable (and lower) F1 values across the continua and primarily vary in the F2 dimension. The acoustic cue variability of the stimuli cannot fully account for the listeners' behavior for the exposure *and* generalization talker as the distributional characteristics are misaligned for such an explanation. There are potentially other cues that may have been missed or altered during resynthesis that listeners are attending to, but it's not readily apparent that is the case. Overall, it seems unlikely that either of the proposed short-term distributional characteristics of the stimuli account for the reduction in /eɪ/ responses at post-test and do not reconcile the generalization results.

9.1.2 Long-Term Distributional Properties

Overall, causal ambiguity may best explain the results of perceptual learning in the /eɪ/ condition. As described above (see also Chapter 2), inference takes place under uncertainty about the true cause of the observed (i.e., perceived) events. Listeners can only infer causality of variation without direct observable evidence of the cause. Listeners therefore engage in inference from a priori knowledge of the distributional properties and underlying causes of variation (causally ambiguous), unless provided with direct disambiguating evidence (causally unambiguous; Liu & Jaeger, 2018). The experiment presented in this chapter assumes that listeners' inference of a single cause will be strong enough to drive behavior. The conjectured single inferred causes are: idiosyncratic tendencies of the talker (the /ʊ/-Biased condition) or the dialect background of the talker (the /eɪ/-Biased condition).

Although those assumptions drove the design of the experiment, it's plausible that listeners have more than one (and possibly competing) hypotheses to draw on during inference, particularly for /eɪ/, which may have more causes of variability compared to a more limited set in /ʊ/. The asymmetry of plausible causes of the two categories may have led /ʊ/ variation to be less causally ambiguous than /eɪ/ where previous experiences may point towards numerous probable causes, including dialect variation, making it maximally ambiguous. As a result, listeners may maintain uncertainty about the true talker-specific characteristics for /eɪ/ because the input they encountered for the talker remains causally ambiguous. While for /ʊ/, the high degree of talker-specificity for this category may have made talker characteristics more prominent as a cause, leading to greater certainty about the talker's pattern.

There are several factors that potentially contribute to the listeners being unlikely to infer the cause of the underlying shift as dialectally driven, or at least rule out other causes. Below, I focus on three potential sources derived from long-term experience that could have maximized listener uncertainty in post-test categorization for /eɪ/ and reduced the plausibility of a dialectal cause for listeners. The explanations all center on listeners' predictions about the speaker and the vowel category based on their long-term experience with American English. Drawing on the discussions across the corpus analyses in Chapters 4-5, these explanations can be linked to listener expectations with regard to the specificity of cross-talker vocalic variation, including typological expectations and relationships among vowel categories. These explanations don't negate dialect as a driving mechanism for perceptual learning, but rather provide a more nuanced perspective of the specificity and additional constraints that may be entailed in listener expectations about dialect variation. However, they do challenge the more holistic model of 'dialect-informative' categories and their status in perceptual learning. The proposed alternative factors below are not necessarily mutually exclusive, and cannot be disentangled under the current results, but provide potential explanations to be pursued in future work. In the sections to follow, I will discuss each of these hypotheses in more detail before turning to the generalization results and a broader conclusion.

First, listeners' experience with /eɪ/ globally may have led them to have a wider range of causal explanations for /eɪ/ variability beyond dialectal variation, resulting in hypotheses of equal or greater strength about the underlying source of the variation. Previous work has argued that

categories that are more widely variable are less perceptually malleable, and therefore demonstrate less category retuning in perceptual learning (Kataoka & Koo, 2017, Stevens et al., 2007). However, such an explanation doesn't explain *why* more variable categories are less prone to retuning; causal ambiguity may be one reason. In relation to the results, listeners' prior experience with /eɪ/ variability may have several known causes, in contrast to /ʊ/ where variability, and its likely causes, are more constrained. That is, /eɪ/ is *maximally* causally ambiguous in the experiment, with listeners' unsure of whether the variation is characteristic of the talker or incidental. Without direct evidence that the novel pattern was caused by dialect variation, listeners remained uncertain about this potential source and all other non-talker-specific causes. In other words, it was not a matter of whether they drew on previous experience, but a matter of *which* previous experiences they drew on as plausible explanations for the experienced perceptual events and how strong the evidence was for any singular inference. As a result, listeners reallocate credibility to the /ʊ/ hypothesis during categorization, reducing /eɪ/ responses at post-test.

Having /eɪ/ be maximally ambiguous points to a potential increase in individual variation across the listeners and whether they perceived the shift as characteristic of the talker. As a brief illustration of this point, take for example the raw categorization data from two listeners in the /eɪ/-Biased condition illustrated in Figure 6.26. The listener on the left demonstrates alignment with the aggregate results of learning, with a reduction in /eɪ/ responses from pre-test to post-test. On the other hand, the listener on the right shows the predicted pattern from exposure, demonstrating more /eɪ/ responses from pre-test to post-test, and the ambiguous step (step 3) shows a greater magnitude of change. A hypothesis is that the listener on the left did not infer the percepts to be characteristic of the talker, demonstrating increased uncertainty at post-test. Contrastingly, the listener on the right appears to have inferred the pattern was characteristic of the talker, as indicated by the shift in responses near the category boundary in the direction of exposure. While these are just two examples, they illustrate the potential range of listener behavior with causally ambiguous percepts. There are additionally listeners who demonstrate minimal change in behavior or increased variability in responses from pre-test to post-test. In contrast, in the /ʊ/-Biased condition listeners generally demonstrate greater magnitude of shifts in learning and more consistently in the direction of exposure. That is, for participants in the /ʊ/-

Biased condition, listeners appear to have more often inferred the pattern is characteristic of the talker.

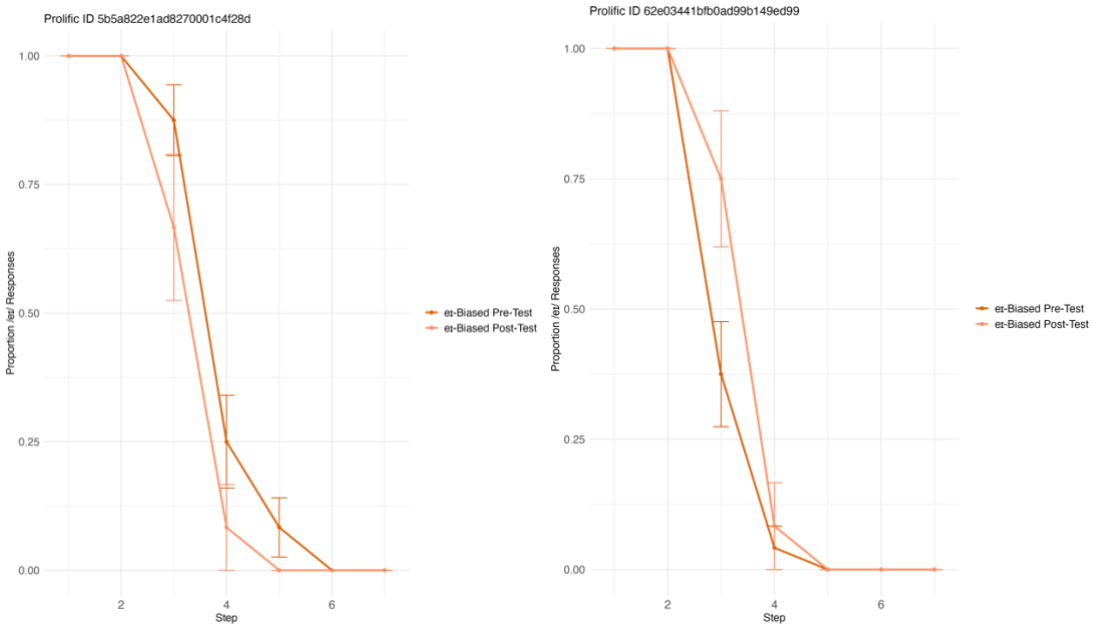

Figure 6.26 Example of participant categorization data from pre-test to post-test in the /eɪ/-Biased condition. The error bars represent the standard error for each participant.

A plausible reason for the individual level variation is that listeners from different regional backgrounds may have variable long-term experiences with /eɪ/ varying across talkers (or varying in different ways). While the corpus data suggest that /eɪ/ varies across all dialects and is predicted by dialect area, there may be more specific expectations about variability within dialects. For example, because /eɪ/ tends to vary more across talkers in the South, generally varying in terms of centralization, we might hypothesize that listeners from the South learn the exposure pattern more readily in contrast to listeners from other regions. However, examining the raw data in Figure 6.27 illustrates that listeners in the South (N = 55) are similar to listeners across other regions (N = 146), demonstrating the same direction of effect with a reduction in /eɪ/ responses at post-test. That is, Southern and non-Southern participants demonstrate the same overall trend and absence of learning the exposure shift (albeit, with the South showing greater reduction of /eɪ/ responses). While listeners' region is not balanced across the dataset, visual data exploration shows similar results across all regions. Thus, the reported regional background of

listeners alone does not appear to account for the individual variation or the aggregate results. Therefore, it's unlikely the causal ambiguity is a function of listeners' experience with dialect-specific variability. At least for /eɪ/ in this experiment, the dialect-specific perspective (see Chapter 4) does not provide strong enough evidence to negate the more dialect-ambiguous hypothesis explored in this chapter.

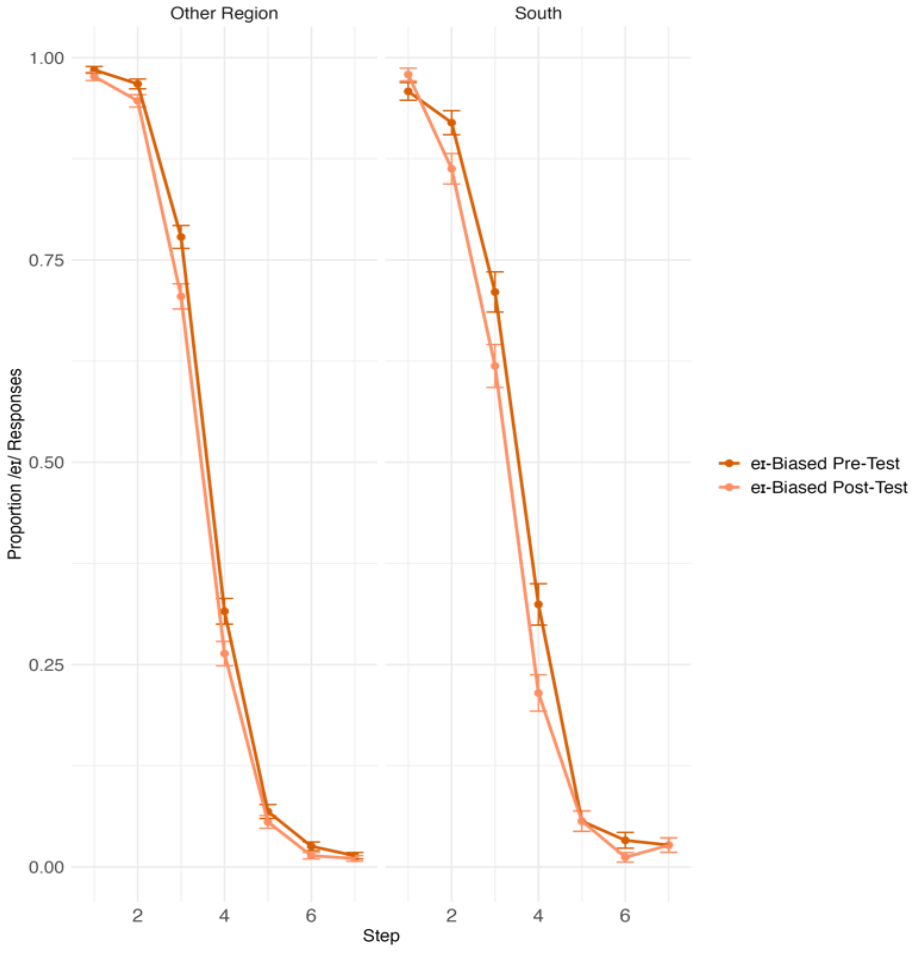

Figure 6.27 Southern listeners compared to non-Southern listeners categorization from pre-test to post-test. Error bars represent standard error.

A second possibility is that listeners made predictions about the vowel inventory of the talker based on her inferred dialect background but after exposure discarded dialect as a plausible cause because there was a mismatch of certain vowel pairs' positions based on prior experience. Previous work offers some support for this hypothesis. For example, Brunellière and Soto-Faraco

(2013) demonstrate that listeners integrate a speaker's accent during word recognition, showing neurological responses indicative of surprisal when processing a word that does not match the surrounding phonological information from the dialect. In the experiment of this chapter, it's plausible that, despite careful consideration to limit the amount and type of other speech heard by the exposure talker, other vowels present during exposure (e.g., /æ/ and /a/) signaled expectations about both the identity of the talker and the production of the critical vowel (i.e., what an /eɪ/ from this talker *should* sound like). The novel shift of /eɪ/ may exhibit a mismatch with listener expectations of the talker's dialect area. If listeners are maintaining multiple potential causes as hypotheses, and talker dialect is a highly probable cause a priori, the mismatch may have caused listeners to reallocate credibility to other potential causes after exposure to other vowels in the system. Thereafter, listeners were more uncertain about whether the pattern was characteristic of the talker and/or a group. Likewise, listeners may have reallocated credibility to /ʊ/ at post-test because prior knowledge about /ʊ/ provided stronger expectations regardless of other details of the vocalic system and/or because there was a greater expectation of /ʊ/ being less variable than /eɪ/, especially after exposure.

While I attempted to control listeners' exposure to other aspects of the speaker's vocalic system, it was unavoidable that listeners received some exposure to additional vowels. The filler vowels /a/ and /æ/ may have incidentally provided listeners with information about the speaker's dialect background and/or triggered expectations about /eɪ/ productions. As observed in the analyses presented in Chapters 4 and 5, as well as much prior sociolinguistic work, both /æ/ and /a/ are critical components of dialect variation across the U.S. While the distributions of /æ/ didn't appear to be informative of dialects broadly in Chapter 4, it likely contributed to evaluation of the speaker's social characteristics regardless. Moreover, /a/ did demonstrate dialect informativity, which likely provided some degree of indication of the talker's dialect background. If listeners were to attribute the cause of variation to the speaker's dialect a priori, the mismatch between expected and observed events may have prompted listeners to become uncertain about attributing the variation to the speaker's identity. Felker et al. (2019) observed a similar pattern whereby Dutch listeners exposed to a novel shift of /ɛ/ towards /ɪ/ resulted in an increase in /ɪ/ responses at post-test, the opposite pattern expected from the exposure. They argue that the increase in /ɪ/ responses may have been a function of an initial perceptual bias towards /ɪ/

and additional exposure to the talker's voice during the exposure, consisting of an interactive game, which influenced how listeners mapped the talker's vowel space. Listeners in the experiment here demonstrate a similar perceptual bias towards /ʊ/ at pre-test, aligning with the explanation given by Felker et al. (2019). Future work would benefit from understanding whether listeners have underlying knowledge about vocalic relationships and whether the presence of other categories in exposure may cause different listener behaviors.

Finally, an alternative is that listeners made predictions about the directionality of variation for certain categories and the shift in exposure did not match their predictions resulting in 'incomplete' information to effectively infer causality. Namely, causal ambiguity may be heightened for listeners because the exposure shift (/eɪ/ to /ʊ/) occurs in a typologically unexpected direction. Babel et al. (2021) found an asymmetry in learning, where listeners are more likely to demonstrate a global relaxation of criteria for phone categorization when exposed to a typologically uncommon shift (voicing of /s/ → /z/) and a targeted adjustment when exposed to a typologically common shift (devoicing of /z/ → /s/). As discussed in Chapter 5, it is typologically uncommon for front vowels to back, but the complementary tendency of back vowels to front occurs systematically across vowel shifts in the U.S. Relatedly, there is a typological tendency for long vowels to raise rather than lower. While these typological patterns speak to historical processes of vowel shifts rather than purely acoustic variability, and it's not uncommon for front vowels to centralize (e.g., /ɛ/ retraction), it's plausible that listeners were unlikely to accept the degree of the shift in /eɪ/ as purely phonetic. Felker et al. (2019) argue that there's an asymmetry in perception of vowels such that peripheral vowels serve as a perceptual anchor, leading listeners to bias towards the more centralized category during phoneme categorization, which was strengthened by the listener's exposure to other vowels from the same speaker. As such, it's further possible that in the present study, the combination of both the typological unexpectedness and exposure to parts of the speaker's vowel system worked in parallel to cause listeners uncertainty and reduction in /eɪ/ responses. Overall, listeners may demonstrate a sensitivity to specific directions of shifts, either through expectations of typological regularities or through perceptual uncertainty driven by peripheral vowels becoming unreliable anchor points. In either case, listeners likely have no prior experience for the pattern

and ‘incomplete’ knowledge to infer a cause. In return, they reallocate credibility to /ʊ/ about which they now have more confidence.

To summarize, the learning results challenge the initial hypothesis that dialect-informative vowels would show evidence of learning in the exposure pattern. I have covered a range, albeit a non-exhaustive list, of potential explanations for the reduction in responses at post-test in the /eɪ/-Biased condition. These explanations challenge the dialect-informative hypothesis that assumes a relatively holistic view of vocalic variability. While we cannot fully disentangle potential explanations, it would suggest that there are other factors that influence how listeners adapt to atypical productions in dialectally conditioned contexts. These range from knowledge about the specificity of dialectally conditioned vocalic variability to a broader understanding of how within-category variability may lead to uncertainty during inference. Additionally, the results for the /ʊ/-Biased condition lend initial support for robust learning to occur for categories that generally show high talker-specific variability but minimal dialect variation.

9.2 Generalization

The hypothesis that cross-talker generalization is more robust (or less constrained) for vowels compared to consonants was not entirely supported by the experiment results. Typically, generalization has been demonstrated through an extension of the learned pattern from exposure. However, in this experiment, there is evidence of listeners generalizing their beliefs about the category following exposure, both in the form of a reduction in /eɪ/ the /eɪ/-Biased condition, and an increase in /ʊ/ in the /ʊ/-Biased condition. Additionally, the results are counter to hypotheses and previous research that show generalization to same-gender pairs but not different-gender pairs. There was no evidence that generalization was constrained to the dialect-informative category (/eɪ/) and talker-specific learning for the talker-informative category (/ʊ/). Additionally, these differences between vowel categories did not correlate with asymmetries in whether generalization occurred in a same or different-gender pair.

These results are surprising given previous work suggests a link between talkers’ acoustic similarity and the degree of generalization (Kleinschmidt, 2019; Kraljic & Samuel, 2006). Similarly, the results are unexpected given past work on cross-talker generalization is

demonstrated via extension of the *learned* pattern of exposure. I will argue, the generalization results in this experiment may be attributed to the perceptual similarity of the talkers' vocalic range of the stimuli, rather than the acoustic similarity. Such an argument is in-line with Reinisch and Holt (2014) who observe a similar pattern in perceptual learning and generalization of /s/. In their study, they altered the degree of match between the perceptual range of the continua from exposure to generalization talkers and observed that generalization occurred when the stimuli were sampled from similar perceptual ranges, but listeners did not generalize if the perceptual range was dissimilar. They also note that the perceptual range may not be the same as the acoustic similarity of the segments.

A similar explanation may elucidate the generalization results of this experiment. As illustrated in Figure 6.5, the acoustic similarity of the segments is greater for the two female talkers' continua (the exposure talker T1_F and generalization talker T2_F), which are similar in overall position in acoustic space and range (from step 1 to step 7). Contrastingly, the male talker's (T3_M) continua occupy a different acoustic position (overall much lower in F1 compared to T1_F and T2_F) and range (greater variability in F2 and minimal variability in F1 compared to T1_F). These patterns would suggest that if listeners relied on acoustic similarity alone, generalization should have been blocked for the male talker but not the female talker. Yet, the opposite pattern occurs. Looking at listeners' categorization behavior across the three talkers in the norming task (Figure 6.4) suggests that, despite acoustic similarity, the categorization functions for T1_F and T2_F are different, where T2_F demonstrates a shallower slope of categorization compared to T1_F. Similarly, T3_M shows similar categorization slope to the female talker despite different acoustic make-up. The difference in the slopes of categorization suggest that while acoustically the two female talkers are more similar, the perceptual range of their continua are different, with T2_F showing less categorical behavior from listeners compared to T1_F and T3_M. Thus, the generalization results suggest that perceptual similarity is a stronger factor in predicting cross-talker generalization for these stimuli than acoustic similarity alone.

In terms of the hypothesized socio-indexical features, it appears that the dichotomy between talker-informative and dialect-informative does not predict listeners generalization behavior. Taking a broader perspective on generalization as listeners extending their beliefs

about a category following exposure, both categories show generalization according to the perceptual similarity of the generalization talkers' continua. While the generalization results contest the original hypotheses, they do not conflict with an ideal adapter account. In fact, this may lend support to the fact that vowels may generally be conducive to greater cross-talker generalization. Similarly, the fact that perceptual similarity drives the generalization behavior is in line with accounts of the model, whereby listeners generalize to the *similar* (i.e., perceptually and/or acoustically similar), despite not aligning with the socio-indexical hypothesis. Lai (2021) suggests a general constraint where adaptation of the exposure talker involves the raw acoustic distributions, whereas cross-talker generalization involves resolving the acoustic values relative to the novel talker's phonological space, involving a degree of abstraction. Future work is required to understand the relationship between perceptual and acoustic similarity.

Finally, we can't rule out the possibility that listeners are inferring an underlying dialectal cause, and indeed such a finding may support this hypothesis as speakers within a dialect area may have perceptual spaces that are more similar. The fact that T2_F is sufficiently dissimilar from T1_F may trigger blocking of generalization as listeners are unlikely to associate the pattern with the same underlying cause. Contrastingly, the increase in uncertainty in the /eɪ/ condition and the learned pattern in /ʊ/ is generalized to the male talker. However, there's no strong disambiguating evidence that listeners extended the vowel patterns based on the link to talker identity or long-term knowledge about different levels of socio-indexical structure. As such, it's also possible that generalization may be a low-level perceptual process that does not entirely rely on top-down higher-order abstractions about links to social structure. Lai (2021) argues that socio-indexical factors may provide gradient constraints on learning and generalization rather than categorical influence on behavior (e.g., blocking). Future work may benefit from integrating explicit talker information as an additional mechanism in understanding the role of socio-indexical factors in perceptual learning and generalization (see e.g., Lai 2021).

10 Conclusion

The learning and generalization results taken together highlight some potential challenges for perceptual learning and generalization and socio-indexical structure in ideal adapter models. Overall, the results do not demonstrate the hypothesized patterns of learning and generalization

predicted by the asymmetrical links to socio-indexical factors. Listeners demonstrated learning and generalization of the talker-informative category, suggesting the underlying cause of the variability was more likely to be inferred as characteristic of the talker. On the other hand, listeners show an increase in bias towards the opposing category in the dialect-informative condition (absence of learning) but show a generalization of this bias to a perceptually similar talker. This finding is puzzling but may speak to the fact that generalization is a low-level process that may not entirely rely on top-down inferences (e.g., Lai 2021), or that vowels overall may generally be conducive cross-talker generalization. The resultant behavior at post-test across the two conditions may largely suggest that these two vowels result in asymmetrical perceptual learning behavior, but it is currently unclear what is driving the asymmetry. It does not appear that a priori beliefs about /eɪ/ being dialectally conditioned are operating here, and if they are it may be that listeners continue to maintain incidental causes of variation equally. Alternatively, other a priori beliefs about the speaker's dialect background or the direction of variation may reduce listeners beliefs about whether the pattern is dialectal or even characteristic of the speaker. Future work should aim to disentangle competing causal models about variability among different vowel contrasts, the directions of shifts, and how relationships among vowels influence perceptual learning.

CHAPTER 7: DISCUSSION & CONCLUSION

1 Introduction

This dissertation examined how vocalic variation is structured within and across talkers and how the systematicity of socially conditioned variation influences listeners' perceptual learning behavior. The background in Chapter 2 provided the foundation of this work and described the complexity of socio-indexical structure, calling for work in speech processing to integrate sociophonetic and sociolinguistic insights into theoretical and computational models of perceptual learning. Therein I outlined several open questions about the theoretical assumptions of socio-indexical structure, stressing the importance of describing dialect areas in more nuanced ways. I highlighted Bayesian models of speech processing which have recently attempted to integrate socio-indexical structure and its function in perceptual processes. Such models make several simplifying assumptions about the nature of socio-indexical structure, including a supposition of a relatively homogenous group depiction, a theoretical question that has been of debate in sociolinguistics. In light of this foundation, I outlined a taxonomy for variability influenced by Guy (1980) that challenges this notion and describes different ways variability may be conditioned by individual and group behavior and their correlation. Given these larger theoretical goals, I will review some of the major findings of this dissertation (Section 2) and then revisit some of the major theoretical implications in light of them (Section 3).

2 Major Findings

Building from this foundation, the first part of this dissertation (Chapters 4-5) examined different analytic scopes of the relationship between dialect areas and individual talkers and prior experience using corpus phonetic approaches. These corpus analytic chapters are meant to probe the complexity of socio-indexical structure and demonstrate how such approaches can inform theoretical and computational models by providing diverse naturalistic data from which to generate testable hypotheses. Chapter 4 extended the ideal adapter model outlined by Kleinschmidt (2019) by simulating different baseline exposures to validate whether the model holds up to more linguistically and socially diverse data. Additionally, this chapter demonstrated

that vowel categories may be asymmetrical in social conditioning, where some categories' distributions are conditioned by dialect areas (categorized as dialect-informative vowels) while others are conditioned by individual talkers and not dialects (talker-informative). Likewise, the chapter directly evaluated whether talkers' distributional patterns align with their dialect areas and demonstrated that for dialect-informative categories, individual talkers within those areas are more regularly patterned in contrast to talker-informative categories where dialect areas do not account for individual distributional patterns. These results provide initial evidence that dialectally conditioned variation results in greater regularity among talkers within a region compared to categories that are not robustly conditioned on dialect areas. It further highlights some of the challenges of defining the analytic levels of socio-indexical factors in theorizing how listeners attend to and evaluate socially conditioned variation. As discussed within the chapter, various analytic levels of socio-indexical factors provide mixed predictions with regard to how specific or generalized listeners a priori beliefs are about socially structured variation and their role in perceptual learning.

Chapter 4 began to challenge the relatively holistic and generalized assumptions in ideal adapter models that all vowel categories have equal likelihood of being socially meaningful and provide listeners with the requisite a priori knowledge to aid in disambiguating sound categories. However, it still assumes that socio-indexical structure operates on an individual category basis in multivariate space, presupposing within category variability of specific cues or relationships among categories is less relevant. Following from this more generalized perspective, Chapter 5 moves towards more specificity in models of socio-indexical structure by drawing on more typical analytic methods of sociophonetics that account for internal structure in the conditioning of social variation. I focused specifically on two components of internal structure, the relationships among vowel categories, represented by acoustic overlap, and variability along specific cue dimensions.

In terms of vocalic relationships, I posited the acoustic overlap of vowel pairs provides listeners with information about the relative boundaries of the category productions which may influence when and how listeners adapt to variation. As one example, I suggest that the acoustic overlap of vowel category pairs in the middle of the vowel space is (slightly) attenuated by considering talker or group identity as an additional dimension in a multivariate space.

Additionally, drawing from previous sociophonetic work, I suggest that some relationships provide stability in the vowel space across talkers. These two facets may provide greater flexibility or constrain listeners' adaptation to variation and influence perceptual categorization. For example, categories where overlap is greatest may result in fuzzy boundaries for listeners making them more malleable for adaption to novel variation. On the other hand, categories with less overlap may demonstrate less flexibility in adaptation and may have sharper category boundaries.

In addition, Chapter 5 demonstrated common patterns of within and across talker variation of specific cue dimensions, such as back vowels demonstrating greater variability along F2 than F1. Relatedly the results also demonstrate that for some categories (like /ɔ/) broad variability is both the result of systematic differences across talkers, within and across regions, and token variability within talkers. While, on the other hand, some categories demonstrate systematic differences between dialects but high regularity of talkers within regions and regularity within talkers. Overall, the results and discussion in Chapters 4 and 5 provide initial empirical foundations for analytic approaches to socio-indexical structure and the a priori beliefs listeners may have about cross-talker variation, internal structure, and its limits.

In the second part of this dissertation (Chapter 6), I provided an example of how large-scale corpus analyses can inform hypothesis generation and testing in perceptual learning experiments. Drawing from the dialect-informative and talker-informative dichotomy outlined in Chapter 4, I selected two vowel categories to test in a lexically guided perceptual learning experiment (/eɪ/ and /ʊ/). I hypothesized that the dialect-specific category (/eɪ/) would promote learning and generalization to novel talkers in a novel shift from /eɪ/ → /ʊ/ (e.g., p[eɪ]stry → p[ʊ]stry). On the other hand, the talker-informative category (/ʊ/) would promote more talker-specific learning and restricted (i.e., same gender pairs) or no generalization to novel talkers in a shift from /ʊ/ → /eɪ/ (e.g., b[ʊ]shes → b[eɪ]shes). These hypotheses were not supported by listener behaviors. Rather, listeners demonstrated a reduction in /eɪ/ responses at post-test for the exposure talker for the dialect-informative category, which I interpreted as an increase in uncertainty in the category. Correspondingly, listeners generalized this uncertainty to the novel male talker but not the female talker. In the /ʊ/ condition, listeners demonstrated learning of the exposure talker's shift and, likewise, generalization to the male talker but not the female talker.

These results suggest that listeners' inferences may not necessarily drawn from a dichotomy between dialect-informative and talker-informative as originally hypothesized. Rather, listeners appear to demonstrate rather complex behavior with generalization that is potentially constrained by perceptual similarity of the talkers and/or stimuli (i.e., perceptual range). The findings of /ʊ/ bring to light that even if a particular category demonstrates greater conditioning on talkers, it does not prevent listeners from generalizing the behavior to other talkers. In other words, while /ʊ/ demonstrates a high degree of talker-specificity in the corpus data, it did not prevent listeners from generalizing the pattern. This finding adds to the growing body of literature that suggests generalization is greater for vowels in general (Kleinschmidt, 2019; Maye et al., 2008; Weatherholtz, 2016), but the asymmetry of effects for these vowels warrants additional research to fully understand the accuracy of such a claim. Additionally, learning may be constrained by other factors including category structure, typological expectations of directionality for vowel shifting (/eɪ/ unlikely to shift towards /ʊ/), or inferences about dialect expectations from surrounding vowel sounds—or some combination of these factors. Descriptions such as ones in Chapter 5 may speak to the experimental results, but future work is necessary to understand the interplay between socio-indexical levels, internal structure, and perceptual learning.

In the following section I will revisit these findings and more specific details as necessary to discuss the theoretical implications of this work. I will begin with a discussion around the nature of socio-indexical structure, focusing on the intersection of within and between-talker variability across dialect areas, and implications for (socio-)phonetics and studies of structured variation. Following this I will consider the implications for listener knowledge and inferential models of perceptual learning. Next, I will revisit some of the major open questions that remain, or arise from, the work in this dissertation and consider the methodological insights and challenges of cross-discipline research, and finally conclude.

3 Theoretical Implications

3.1 Characterization of Socio-Indexical Structure: Individuals and Groups

By examining the distributional properties of vowels, this dissertation advances our understanding of vocalic variation across regional dialects and makes contributions to sociophonetic theories with regard to the nature of individuals within their broader dialect areas. Importantly, I have illustrated that individuals' distributional properties largely mirror their dialect areas, particularly when dialects are predictive of cue distributions across American English (Chapter 4-5). Such a finding provides additional support for theories in sociophonetics that suggest talkers reproduce their dialect areas' patterns despite a wide range of talker variation within dialects (see e.g., Guy, 1980; Oushiro, 2016). The data further validate recent observations of structured variation in phonetics, whereby dialect areas may be differentiated by their central tendency for a given cue to contrast mapping, but individuals within those areas are regularly patterned along the same lines (Sonderegger et al., 2020). However, it has also highlighted that socio-indexically structured variation may promote different distributional shapes and structure to variability. In light of these patterns, I will revisit the Taxonomy of Variability detailed in Chapter 2 (and again in Table 7.1 below) with these findings in mind, followed by a discussion of implications for sociophonetics.

Across analyses in Chapters 4-5, several patterns emerged that fall within each of the types of variability listed in the taxonomy (Table 7.1 below). Type 1 may best characterize /u/ for these data, where there is a broad range of variability at the broader population level (i.e., American English), and it seems to, by and large, remain at the dialect and talker levels. While there is evidence that /u/ varies by dialect area, it largely shows the same pattern of token level variability from talkers to dialect areas which may be indicative of segmental context-induced variation (i.e., fronting before coronals). Type 2a appears to largely describe /eɪ/, where means distinguish dialect areas and talkers within dialect areas are highly regular, and the spread of the distributions is relatively similar across all socio-indexical levels. Type 3 may be best captured by /ʊ/, where variability is not conditioned by dialect areas and individuals appear to be more idiosyncratic in their productions (talker-specific). Finally, Type 4 may best describe /ɔ/

variation, where dialect groups condition variability but there appears to be systematic differences across talkers within regions and pervasive token variability within talkers.

Table 7.1 Taxonomy of variability, adapted from Guy's (1980) taxonomy of variation.

Individuals	Groups	
	Similar	Different
Similar	1. uniform force; same mean and same variance	2a. Different means; same variance OR 2b. different means and different variance (Social or geographic dialects)
Different	3. Individuals with different means and/or low dispersion (Individually stratified linguistic variation)	4. Combinations of 2 and 3, or true free variation

These fine-grained phonetic patterns provide general insights into expectations for what types of variation may be mirrored from the community to the individual. While it's unclear what the different sources are that lead to this variability, the starting point may be important for understanding the range of variability that exists within a community, and determining how they relate to descriptions of community patterns. Additionally, understanding these different types may inform whether individual variation aligns with community variation along the same axes; that is, the extent to which within-talker variation mirrors patterns of social stratification. For example, Bell (1984: 151) posits a style (i.e., formality) axiom that "variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the 'social' dimension". By looking at distributional properties across and within talkers, researchers may elucidate whether such an axiom holds, and the related role of linguistic factors in relation to social dimensions (see e.g., Preston, 1991). However, much of the data in this dissertation do not encompass the type of individual variation that would provide the most insight, thus leaving open questions in this regard. Other more typical sociophonetic studies of individual variation are apt to shed light on the more socially sensitive and fine-

grained indexical variation across the distributional space (see e.g., Podesva, 2011; Van Hofwegen, 2017).

Beyond sociolinguistics, work examining variability in speech has attempted to categorize different sources and types of variation for some time. Recent work examining structured variation in phonetics has described a continuum of individual and community norms, where on one end talkers may be maximally agentive and on the other end maximally reflective of their communities (Chodroff & Wilson, 2017). The work in this dissertation demonstrates that there are a range of potential points along that continuum and, where a given contrast, talker, and community falls along the continuum, may be a function of the unit (e.g., means, distributions, etc.) and scope (e.g., grouping structure), and time course examined (see also Tamminga & Wade, 2022).

In addition to these points, the data in this dissertation suggest that it may vary depending on the source and type of variability. Elman and McClelland (1984, 1986), for example, describe "lawful" sources of variation, those that arise from regularly governed phonological or segmental context, in contrast to variability that is less discrete or more "random" including that of cross-talker and token-level variability. Comparably, Cohn and Renwick (2021) argue that the phonology-phonetics relationship can be summarized under contrasting dimensions of variability divided into 'systematic' and 'sporadic' and dimensions of gradience from 'gradient' to 'categorical', of which the socio-indexical level of analysis (i.e., social group or talker) intersect through this space. Early work in sociolinguistics can be summarized as describing systematic categorical variation, in that impressionistic coding was predominate methodological measure of variation, and social factors were seen as an analogous extension of context in phonology. By examining the gradient and continuous acoustic distributions, we may be capturing systematic 'lawful' variation (e.g., phonological context) or sporadic 'unlawful' variability which may be defined by other more microsocial organizations, idiosyncrasies, physiological differences, or statistical noise. Note, of course, these are not actually unlawful, but require different analytic lenses to uncover the full regularity, aside from true statistical noise. For example, the distributional make-up of /ɔ/ illustrates the broader population variability reflects both systematic differences between talkers (e.g., differences in means across regions) as well as variability within talkers. Whether such patterns may be the result of the "lawful" variation or sporadic

token variability still remains an open question (see Quam & Creel, 2021 for similar discussion in child language acquisition). Additionally, in the case of /ɔ/, it's unclear from these data whether ongoing merger influences the token variability within speakers as a function of, for example, lexical variation (see e.g., Warren & Hay, 2006). Nonetheless, this dissertation illustrates the benefits of drawing on large corpora as it allows for the identification of phonetic patterns without assumptions about the mapping of such variability a priori to further refine theories of socio-indexical variation.

3.2 Listener Knowledge & Previous Experience

There is mounting evidence across linguistics that listeners represent variability in speech and use it to guide speech processing. The work in this dissertation speaks to how variable experiences with American English may converge to reveal patterns of socially conditioned variability for a given category. In each of the simulations in Chapter 4, the findings revealed notable patterns that speak to potential listener representations of variability. In particular, looking across vowel categories, it demonstrated that informativity was highest for some regional varieties, like the South and North. This finding validates listener perceptions of the same regions as being highly salient across vowel categories. In addition, it highlighted several categories that distinguish dialect areas with respect to a broad description of American English, and from one another, as in the case of using the West as the reference distribution.

The finding that informativity was not evident at the most salient categories associated with vowel shifts or socio-indexical style poses interesting questions into how talker regularity may provide bottom-up information about social groups. This finding speaks to previous work in sociophonetic perception studies, where listeners are able to categorize and evaluate talkers even when salient cues are absent (Clopper & Pisoni, 2004a-c, 2007; Gunter et al., 2020). The distributional patterns outlined in Chapter 4 and Chapter 5 illustrate that some categories' cue distributions may be more readily partitioned in acoustic space when conditioned on dialect areas by being marked with a high degree of regularity of talkers within groups (e.g., /eɪ/). However, the data illustrated this may not necessarily be reflective of categories that have been demonstrated in previous research to be highly identifiable across regional dialects (e.g., /æ/). While there may be several reasons for this, one potential explanation is that some categories

distinguish dialects from one another (i.e., /æ/ in the NCS vs. /æ/ in LMBS) but may be unlikely to be highly distinctive when examined in the aggregate of American English more broadly across a flat distribution of tokens.

Indeed, Chapter 4 demonstrated that when using the West as the reference group, categories like /æ/ and /a/ emerge as more strongly conditioned by dialect compared to the aggregate American English simulation. Some categories are best represented in terms of differences between individual dialects rather than being divergent from an aggregate form of American English. This is both a methodological and theoretical interest, as computational models evaluating listener perspectives may wish to consider theoretical underpinnings to these different reference points. Such findings may be integrated with existing theories of cognition to interrogate how much or what kind of exposure may be necessary for encoding and representing such variability (see Docherty & Foulkes, 2014 for additional discussion). This finding warrants continuing work examining the relative weight that listener's give to their own dialect backgrounds versus the generalized knowledge they have about (American) English more broadly and under what constraints and perceptual tasks listeners draw on these different experiences.

3.3 Listener Inferences & Perceptual Learning

A key contribution of this dissertation is examining the relationship between what talkers do in production to what listeners potentially infer and do in perception, focusing specifically on perceptually learning. Previous work has posited that listeners act as ideal observers by learning the parameters of cue distributions across talkers to infer the likely cause of perceptual experiences. This dissertation tested one facet of ideal adapter models by asking whether listeners' perceptual learning behavior can be predicted by the regularity of a contrast being conditioned on group and individual identity. The experiment in Chapter 6 provides some initial evidence of belief updating models, albeit in less straightforward ways than initially hypothesized.

3.3.1 Learning

The learning behavior exhibited by participants across the two categories may have demonstrated different belief updating, where the /ʊ/-Biased condition demonstrates the learned pattern and /eɪ/-Biased condition demonstrates a narrowing of the category. There are several hypothesized reasons why this occurred, as outlined in Chapter 6. Listener behavior may nonetheless point to the integration of the short-term distributions with long-term previous experience and updated beliefs about category structure. I hypothesized in Chapter 6 that listeners' updated beliefs were reflected in greater uncertainty during post-test categorization and reallocating credibility to the /ʊ/ category. Correspondingly, post-test results show greater magnitude of a shift in the /ʊ/ condition, providing credibility to the fact that listeners adapted to the exposure shift for this category. I suggest that /ʊ/ may have been more readily inferred as characteristic of the talker given its narrower range of variability and regularity within a given talker compared to /eɪ/.

In terms of broader theoretical implications, the results of this experiment challenge current descriptions of Bayesian inference relative to socio-indexical structure. While the indicated levels of socio-indexical structure did not appear to be the primary determinants of listener behavior, it does not rule out that they were still involved. As discussed in Chapter 6, there may be other factors of prior experience that listeners track and make use of in socio-indexical inferences for perceptual learning. In addition, a challenge I have not addressed thus far arises from that task itself and the multi-dimensional nature of vocalic variability. Previous work has primarily examined consonants in perceptual learning, whereby the manipulations of individual cues may prove to be a simpler task for listeners to accomplish. On the other hand, vowels are fluid and dynamic and may demonstrate a wider range of variability in everyday speech.

As a result, listeners may respond to such variability in perceptual learning tasks differently than consonant variability. For example, listeners may respond to increasing variability and uncertainty with biased heuristics that are more likely to minimize prediction error in future contexts. For the case of /eɪ/, this may appear as a narrowing of the category structure. Given that listeners may experience variability more often for vowels across talkers and contexts, listeners may employ heuristics to avoid increasing an already variable category

without sufficient evidence. In other terms, listeners may avoid over-fitting to the new experience and opt to bias towards underfitting to the experienced short-term distributions that are outside of the range of previous experience to avoid errors in the future. Such a mechanism may also explain, in part, the generalization of such uncertainty to the perceptually similar talker, as listeners may opt for the same bias without additional evidence. Overall, future work should aim to understand how listeners may employ different strategies especially when faced with complex inferential tasks and multi-dimensional contrasts like vowels (see e.g., Tahra et al., 2022 for a recent discussion in visual perception).

3.3.2 Generalization

The generalization behavior of listeners across the two categories further supports the fact that listeners more globally updated their beliefs in response to the exposure. Listeners in the /eɪ/ conditioned did not restrict their beliefs to the exposure talker, but rather extended it to a perceptually similar novel talker. Similarly, listeners extended the learned shift of /ʊ/ to the same perceptually similar talker. This pattern broadly refutes the claim that spectral information alone is enough to block generalization (Kraljic & Samuel, 2006) and that vowels should demonstrate same-gender generalization patterns as a function of acoustic similarity or higher order socio-indexical structure alone (Kleinschmidt, 2019). Rather, this finding provides evidence that generalization may be a function of perceptual similarity and may encompass a generalization of beliefs more broadly rather than the direction of exposure.

This finding complicates our understanding of generalization in perceptual learning and warrants additional research to understand the interplay between acoustic similarity and perceptual similarity. This aligns with Reinisch and Holt (2014) who demonstrated that manipulating the perceptual similarity of the generalization talkers' test continua to span a similar perceptual range is predictive of generalization behavior regardless of the gender of the talker. Making it less about the perceptual similarity of the voices and more about the similarity of generalization stimuli to exposure items. Recent work by Lai (2021) calls attention to this issue and describes an acoustics-phonology mismatch constraint. Lai (2021) suggests that the adaptation to a talker involves learning the raw acoustic distributions of the category, while the generalization of perceptual learning involves evaluating the relative differences between the

targets of different talkers in phonological space. Rather than the acoustic values overlapping between the target segments of the exposure and generalization talkers, listeners are likely to generalize when the perceptual targets of the phonological space are similar. Such a mechanism may be at work in the experiment in Chapter 6 where the acoustics are more aligned for the novel female talker, but the perceptual phonological space is not. Future work should aim to understand how perceptual similarity and acoustic similarity differ, and whether such a distinction does in fact constrain generalization.

These findings however do not directly refute the possibility that listeners use higher-order socio-indexical structure during inferential processing. It is possible that the perceptual similarity of the test items was enough to induce perceptual learning or there are some low-level constraints of acoustic or perceptual similarity for generalization to occur. In such cases, higher-order socio-indexical structure might operate at a different time course for generalization or be mediated by other cognitive factors such as attention or visual information. Some work has argued that socio-indexical effects may emerge later in processing (McLennan & Luce, 2005). Relatedly, Theodore et al. (2015b) have demonstrated that effects of talker identity in speech processing are mediated by attention allocated to talker identity rather than the time course of processing or difficulty of the task. In a perceptual learning study, Lai (2021) suggests that talker identity provides gradient constraints on listeners' generalization behavior as opposed to categorical blocking (or promotion) of generalization. Lai (2021) demonstrated some evidence that providing visual cues to talker identity attenuates the degree of generalization of a shift in atypical stop productions from a novel talker but only weakly for sibilants. Future work should aim to disentangle how and when higher order socio-indexical knowledge comes to bear on listeners' perceptual learning and generalization behavior adding to the broader debates of the role of socio-indexical factors in speech processing.

3.4 What are Listeners Tracking?

A comprehensive description of the nature of phonetically cued variation that is tracked by listeners still remains an open question. Socio-indexical structure is undoubtedly a multidimensional problem that various disciplines and scholars have wrestled with, ranging from a more social focused understanding (e.g., speech communities, social networks, etc.) to the

more fine-grained phonetic understanding (e.g., structured variation). The findings of this dissertation have highlighted how current conceptualizations of socio-indexical structure may lack perspectives of both the group structure (i.e., individuals vs. social groups) and the internal regularity of socially conditioned variation. The results of the experimental chapter are not easily explained given a model where talker and group specific input is operationalized as the raw cue distributions over a given cue to category mapping.

As one example of the relationship between internal principles and external social factors, Chapter 5 probed how certain vowel pairs may provide stability or instability across talkers. In particular, some vowel pairs maintain the same degree of acoustic separation across talkers and dialect areas (i.e., /a/-/æ/), while other category pairs demonstrate attenuation of acoustic overlap when considering socio-indexical factors. As such, it's apparent that the relationships between vowel categories are shaped both by social factors and internal principles of variation within the vowel space. I emphasize the quantitative analysis of socio-indexical variability therein is not meant to stand as a proxy for a cognitive representation, but rather provides descriptive statistics for distributional patterns across socio-indexical levels and vowel pairs. As such, the patterns observed prompt interesting questions about the degree of specificity that may constrain listeners' perceptual learning and generalization behavior. Future work should aim to validate the descriptions therein and further elucidate to what extent perceptual beliefs are updated by tracking the cue distribution of the single category or the relative locations and boundaries to other category instances.

Relatedly, it's unclear to what extent listeners rely on the internal category structure when tracking cross-talker variation. As an example, the wider range of variability in /ɔ/ was demonstrated across and within talkers. When learning talker-specific systems, do listeners behave differently for categories where the within-talker variability is on par with cross-talker variability? Current work examining category dispersion would suggest not and that listeners may be more likely to demonstrate uncertainty in learning (Clayards et al., 2008). These different axes of internal category structure alongside socio-indexical factors should be the aim of future work to continue to refine.

This dissertation suggests that current inferential models may not adequately capture the relevant level of socio-indexical structure that is used for speech adaptation and generalization

behavior. The results of the experiment (Chapter 6) for example suggest that listeners may not be sensitive to the broad distinction of dialect-informative and talker-informative, at least with regard to the vowel categories in the experiment. The results provide weak evidence that listeners are likely to generalize updated beliefs about categories to perceptually similar talkers. However, it's unclear whether prior experience or inferential processes inform listeners' behavior or whether socio-indexical structure can adequately account for the generalization patterns. As outlined in Chapter 6, it's plausible that listeners' behavior was driven by more fine-grained knowledge of socio-indexical structure among vocalic relationships or typological tendencies of variation. Though, it's possible that the hypothesized socio-indexical level is incorrect or incomplete and is not the dominate influence in perceptual learning and generalization behavior. Given the experimental results and the corpus analyses in Chapters 4-5, it remains an open question what the relative weighting is of individual norms against the community or group norms. Additionally, the dichotomy between individuals and dialects may not be representative of competing socio-indexical sources at all times, as individual talkers' variable tendencies may be reflective of socially meaningful variation within the community at large. Future work should aim to explicate when listeners are likely to draw on various levels of socio-indexical structure and whether the researcher-imposed analytic levels align with listeners' beliefs about socially conditioned variation.

At the intersection of these discussions is a question about whether socio-indexical structure may inform different types of adaptation behavior. One potential factor may be the nature of (atypical) productions listeners are exposed to. Theodore et al. (2015) suggests that the ways in which listeners adjust speech representations to individual talkers may depend on the nature of the target productions. Specifically, Theodore et al. (2015) demonstrate that listeners reorganize the perceptual space of the category itself while keeping the boundaries between categories intact when exposed to unambiguous productions. Alternatively, ambiguous productions may be more likely to induce category boundary shifts as a function of necessity in resolving the ambiguity. It remains unclear whether such ambiguous productions equally promote reorganization of the internal category structure or whether it is in fact restricted to the boundary. Theodore et al. (2015) suggest that unambiguous productions are more aligned with variation induced by talker identity such as dialect or accent, while ambiguous items may be

indicative of idiosyncratic productions. Such a pattern speaks to the results in Chapter 6, as ambiguous categories may be more challenging among vowel categories where productions are more fluid and dynamic compared to consonants, yielding fuzzier boundaries. One potential avenue to examine this question is to examine listeners' shift in internal category structure which may aid in evaluating whether listeners identify tokens as better/worse exemplars after exposure to talkers with atypical productions.

A final point to consider is whether the operationalization of socio-indexical structure over the raw cue distributions is most reflective of listeners' knowledge. It remains an open question whether listeners generalize cross-talker expectations on the basis of raw cue distributions or whether their experiences and knowledge are more fine-grained and contextually bound. This dissertation did not address the intricacies of contextually driven variation, either in terms of socially contextualized or internal phonological context. Recent work has discussed these intersections, with some arguing that talker-specific learning is reflective of individual speech patterns that are not (phonologically) contextually bound (Kraljic et al., 2008; Kleinschmidt, 2019), while others demonstrating within-talker variation is also learned (Idemaru & Vaughn, 2020). On the other hand, learning the social context-induced variation has been suggested to operate under Bayesian mechanisms provided the contexts are sufficiently informative of cue distributions (Kleinschmidt, 2019). Thus, it's unclear when and if listeners track more fine-grained internal linguistic context alongside social factors in conditioning variation. This brings me to my next broader input, the challenges of crossing disciplines and the assumptions therein.

3.5 Crossing Disciplines: Methodological & Practical Implications

The above discussion broadly speaks to some of the theoretical challenges of bridging sociophonetics and psycholinguistic frameworks. The core assumptions of the fields may lead to challenges in integrating insights to and from either domain. By applying generalizations of dialect variation into speech processing frameworks we may miss valuable facets of socially conditioned variation (e.g., internal conditioning). These challenges may largely stem from the differences in the interests between more phonetic approaches which tend to focus on the cognitive and physiological factors of speech, and sociolinguistic approaches which focus on the

social and phonological factors in speech. Of course, as sociophonetics continues to grow, so do the theoretical goals, making these points necessary to address.

Crossing disciplines still poses challenges despite the sharing of theoretical goals and methodological advancements. As described in Chapter 2, much sociolinguistic work defines the unit of analysis as the central tendency (e.g., mean) of a particular category or categories and given the internal conditioning factors. The focus on *variation* may largely identify abstract patterns of variation across a speech community or social group (see e.g., Eckert & Labov, 2017) and relatively categorical forms induced by social and linguistic context. While on the other hand, recent work in speech processing has examined *variability* as broadly stochastic fluctuations of the acoustic signal, which may encompass a broader range of sources and less discrete partitioning of the problem space. To some extent such differences functionally separate pursuits in the respective fields to examine different analytic scopes, with speech processing approaches at the individual level and sociophonetic approaches at the group level, although, this is increasingly changing (see e.g., Sonderegger et al., 2020, Tanner et al., 2020). Divergence in the unit of analysis between the fields prompts questions about how researchers apply understanding of dialect variation from sociophonetic studies largely focusing on group-means to theoretical and computational models that encompass the entire distributional space.

This dissertation speaks to this gap by intersecting methodological aspects of sociophonetics and speech processing given the same unit of analysis (distributional properties of vowels) and under different analytic scopes (individuals and groups). This work has highlighted some broader questions about how dialect areas are distinguished from one another (e.g., is it in central tendency alone?) and whether individual talkers can be seen to mirror community patterns if given enough data (see also Cohn & Renwick, 2021; Tamminga & Wade, 2022). While the emphasis here was speech processing, such perspectives afford insight into a wide range of interests in sociophonetics. For example, examining distributions across individuals and groups would offer insights into theorizing whether sound change is gradient or categorical (see e.g., Fruehwald, 2013, 2017 for such discussion). Given the continuing development of large-scale speech datasets, such an examination is becoming more feasible. As such, future work in sociophonetics may benefit from looking at distributional properties more closely and integrating these insights into current theoretical frameworks.

Theories of speech processing may equally benefit from probing the assumptions and deviations in methods that draw on insights from sociophonetic work. As I have highlighted above, one particular area that sociophonetics often highlights may be to consider how and when internal constraints intersect with socio-indexical factors to guide perceptual learning. And, as I have demonstrated in Chapters 4-5, methods calling on more hierarchical organizations of social groups may further develop theoretic perspectives of socio-indexicality in speech processing. Finally, examining more naturalistic speech and (socio-)linguistically diverse datasets may provide additional insight into the range of variability from which listeners learn. Overall, both fields benefit from the integration of both methodological and theoretical perspectives to generate and test hypotheses about the nature of socio-indexical structure in both production and perception.

As I have attempted to bridge some of these gaps, it has become clear there are several methodological challenges to integrating vowels in perceptual learning frameworks. As may have become apparent in the different components of this dissertation, examining vowels at the intersection of production and perceptual learning posed several obstacles that highlight both their value and their complications. These obstacles are most evident in the lexically guided perceptual learning experiment. Vowel categories provide a great deal of complexity to experimental design in these paradigms due to the multivariate cues, the fluidity of category boundaries, and the prevalence of vowels in the phonological inventory and composition of lexical items. While previous studies that have examined stops or fricatives are able to control the presence and absence of additional input from similar contrasts (e.g., /s/ and /ʃ/), it's impossible to avoid the presence of additional vowel categories during exposure. Relatedly, the multivariate nature of vocalic cues makes it increasingly challenging to identify a single dimension upon which listeners base their perceptual retuning and for the creation of stimuli. As a result, these facets add additional confounds to the experiment that prove difficult to overcome. While I attempted to mitigate these methodological issues, there is no perfect solution. These points alone have elucidated why vowel categories may be under studied in lexically guided perceptual learning. However, these same complexities demonstrate why vowels add a great deal of value and insight into the limitations of current models and methods in perceptual learning and are necessary to study.

Relatedly, using corpora to investigate the multidimensional space of socio-indexical structure provides several challenges. With large corpora, researchers are afforded the ability to look across speakers and gain insights into a variety of phenomena simultaneously and validate phonetic patterns across a diverse range of speakers. However, corpora still have limitations, in part stemming from collection methods, such as limited and variable metadata about the speakers. Most critically, corpora are typically curated on the basis of eliciting natural speech, which is unbalanced in the representation of the phenomena researchers may wish to examine. This poses an issue as researchers attempt to narrow data to specific contexts or balance tokens by talkers, from which large datasets become increasingly small. These challenges continue to call for a multi-method and iterative process to generate and test hypotheses, drawing on corpora, laboratory speech, and perception experiments to gain a more holistic understanding of speech production and speech processing. Thus, while there are inherent hurdles to overcome in interdisciplinary research, it adds great value to linguistic theory to do so.

4 Conclusion

This dissertation advances empirical foundations to socio-indexical structure as it pertains to speech processing by exploring the relationship between variability in speech production and perceptual learning. By examining a large-scale, diverse dataset of American English, this dissertation simulates a broader range of experiences with talkers affording critical data for the generation of testable hypotheses for listener behavior. The analyses therein speak to longstanding debates within sociolinguistics about the systematic aspects of group and individual behavior and the status of individuals with regard to community norms. Informed by these analyses, the inclusion of a perceptual learning experiment of vowels provided valuable theoretical insights for the interplay between socio-indexical structure in production and perception. The results of this dissertation highlight the complexity and challenges associated with the perceptual learning of vowels. Overall, this dissertation bridges the gap between current theories of speech processing and sociophonetic theories of socio-indexical structure to identify analytic and theoretic assumptions about the nature of socially conditioned phonetic variation.

APPENDIX A: EXPERIMENT STIMULI

Non-Words:

galast	robble	kabbath
tanger	gaastage	maddle
lankwatt	zottest	pamon
fanker	pactic	safurd
kashful	laarfer	skobar
dathram	raaffer	shapas
faadas	daagic	plashes
famel	dagnum	sloked
bandle	zartyr	slarter
catob	karvel	saaler
fapten	maspers	doften
chalter	maller	folace
caaptal	magress	ronic
saallege	maafel	staagger
paancert	bodern	zalon
laanscious	naadel	tanka
faaper	gackage	paarget
prackers	farcel	parnish
dasser	daardon	tragel
rances	passak	valap
baacter	pasern	vaalka
fabtest	plafna	saffle
gatten	gaacket	laarget
prasile	ronder	salot
lasket	praakise	drolick
paadess	sather	zatterns
paasip	dattle	paller
narder	dobin	grazzle
pavoc	motten	

Filler Words:

blackness	cancel	columns
blackout	captain	combat
bladder	carrots	comet
blanket	cashews	contract
blossom	casket	convoy
bonfire	catcher	copper
bother	chapter	cottage
bottled	cobbler	dashboard

doctors
dolphin
fashion
flannel
fondness
fossil
gallon
garlic
gather
goggles
grammar
gravel
hammer
hammock
happen
harvest
honest
hostile
jackpot
karma
llama
lobster
locket
magnet
market
matter
monster
packet
padded
paddle
passes
pocket
pollen
posture
raptor
saddest
scholar
scratches
soccer
socket
soften
spotlight
tractor
ratchet
sandal

Critical Items:

/eɪ/-Biased Condition

blazers
blazing
canine
haystack
hazing
majors
maker
maple
nations
native
neighbor
pastry
playground
rainbow
raven
razor
reindeer
tasted
trading
waking

/ʊ/-Biased Condition

bushes
butcher
cookies
cooking (3A)
crooked
footage
football
footpath
goodies (1B)
hooking (3A)
lookout
pudding
pusher
rookie
sugar
woman
wooden
woodwork
woody (3A)
goodness

APPENDIX B: STIMULI ELICITATION MATERIALS

The original recording of participants also included items for a different experiment not presented in this dissertation, thus the stimuli list below encompasses more items than used in the experiment.

Reading Passage (Fridland & Kendall, 2022)

Some mornings in the summertime, when the sky is fair and the lawn covered in dew, the good Duke Post and his wife Peg walk down to the brook by their house. There, beside the trees, is their favorite place to sit, talk and sip coffee.

Her father, Don, and his dog, Bookie, often stop by to chat while their children, Betty and Kate, toss off their shoes and leap headfirst into the deep brook. It makes Peg feel like a kid again to watch them dive, shout and slosh around in the water and swing off the old black tire tied to the oak tree.

One hot hazy, dull afternoon, she gave a call to their friends Pam and Ben Powder, inviting them over for supper. On the way, their truck got stuck in the mud and they showed up an hour late, for which they caught a good deal of teasing.

But soon the crowd was having fun and the good hosts put out tuna fish sandwiches, hot dogs, a big pot of bean soup and beer bread. When they were done eating, it was a sin that no one had saved room for Peg's tasty spice cake that was yet to come.

After supper, Duke, Ben and his pal Bill went out on Dukes inflatable boat. Unfortunately, the sky got grey and started to pour rain. Bill lost his footing on the slick bank and fell in the water. After ten minutes he finally got into the boat. Once back on shore, the sudden weather shift sent everyone home, and the party was over.

Word List

bake
beyklet
beykworm
beyshes
beytcher
bicklet
bickworm
bikcase
bikstore
bishels
bishes
bisim
blazers
blazing
bloozers
bloozing
book
bookcase
booklet
bookstore
bookworm
booshels
booshes
bootcher
bosom
breakage
breaking
brookage
brooking
bukcase
booklet
bookstore
bookworm

bushels
bushes
busom
butcher
cabled
cake
canine
ceykies
ceyking
cishion
coobled
cook
cookies
cooking
coonine
cooshion
could
creyked
cricked
crooked
cruked
feytage
feytball
feytpath
fit
fitage
fitball
fitpath
fitsteps
foot
footage
football
footpath
footsteps
futage
futball
futpath
futsteps
geydies
geydness
gidness
goodies
goodness
gudness
haystack
haywire
hazing
hazy
heyking
hooking
hoostack
hoowire
hoozing
hoozy
kick
kid
leykout
lickout
lookout
lukout
majors
maker
maple
moojors
mooker
moople
nations
native
neighbor
noober
nootions
nootive
pasting
pastry
peydding
peysher
pisher
pit
playground
plooground
poosher
poosting
poostry
pudding
pusher
put
rainbow
raincoat
raven
razor
reykie
rookie
roonbow
rooncoat
roondeer
rooven
roozer
shake
sheygar
shigar
shigared
shooed
shoogar
shook
should
shugared
sit
skater
skooter
soot
stewed
stood
sugar
sugared
suit
tasted
toosted
trading
trooding
waking
weyden
weydwork
weydy
weyman
weyman
widden
widwork
woman
wooden
woodwork
woody
wooking
wuden
wudwork
wuman

REFERENCES CITED

Abrego-Collier, C., Grove, J., Sonderegger, M., & Alan, C. (2011). Effects of Speaker Evaluation on Phonetic Convergence. *Proceedings of the 17th International Congress of Phonetic Sciences*, 192–195.

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520.

Ahn, E., & Chodroff, E. (2022). Voxcommunis: A corpus for cross-linguistic phonetic analysis. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 5286–5294.

Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. *The Journal of the Acoustical society of America*, *51*(2B), 648–651.

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *115*(6), 3171–3183. https://doi.org/10.1121/1.1701898

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*(1), 544–552.

Apfelbaum, K. S., Bullock-Rest, N., Rhone, A. E., Jongman, A., & McMurray, B. (2014). Contingent categorisation in speech perception. *Language, Cognition and Neuroscience*, *29*(9), 1070–1082. https://doi.org/10.1080/01690965.2013.824995

Arnold, L. (2015). Multiple mergers: Production and perception of three pre-/l/mergers in Youngstown, Ohio. *University of Pennsylvania Working Papers in Linguistics*, *21*(2), 2.

Babel, M.. (2009). *Phonetic and social selectivity in speech accommodation*. [Doctoral dissertation, University of California, Berkeley].

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189.

Babel, M., Johnson, K., & Sen, C. (2021). Asymmetries in Perceptual Adjustments to Non-Canonical Pronunciations. *Language and Communicative Disorders 12*(1). https://doi.org/10.16995/labphon.6442

Babel, M., McAuliffe, M., & Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00653

Babel, M., Senior, B., & Bishop, S. (2019). Do social preferences matter in lexical retuning? *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 4. https://doi.org/10.5334/labphon.133

Babel, M., McAuliffe, M., Norton, C., Senior, B., & Vaughn, C. (2019). The Goldilocks Zone of Perceptual Learning. *Phonetica*, *76*(2–3), 179–200. https://doi.org/10.1159/000494929

Baese-Berk, M. M. (2010). *An examination of the relationship between speech perception and production* [Doctoral dissertation, Northwestern University].

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, *133*(3), EL174–EL180.

Bailey, C.-J. N. (1973). *Variation and linguistic theory.* Center for Applied Linguistics, Arlington, VA.

Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, *23*(3), 347–374. https://doi.org/10.1017/S0954394511000135

Bayley, R., & Langman, J. (2004). Variation in the group and the individual: Evidence from second language acquisition. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *42*(4), 303–318.

Becker, K. (Ed.). (2019). The low-back-merger shift: Uniting the Canadian vowel shift, the California vowel shift, and short front vowel shifts across North America. *Publication of the American Dialect Society*, *104*(1), 1–220.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204.

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, *11*(2), 65–85.

Benor, S. B. (2008). Towards a New Understanding of Jewish Language in the Twenty-First Century. *Religion Compass*, *2*(6), 1062–1080.

Bent, T., & Holt, R. F. (2017). Representation of speech variability. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4), e1434. https://doi.org/10.1002/wcs.1434

Bickerton, D. (1975). *Dynamics of a creole system.* Cambridge: University Press.

Bigham, D. S. (2010). *Correlation of the Low-Back Vowel Merger and TRAP-Retraction*. *University of Pennsylvania Working Papers in Linguistics*, *15*(2), 21–31.

Boberg, C. (2005). The Canadian Shift in Montreal. *Language Variation and Change*, *17*(2), 133–154.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. *Version 6.0.37.* Retrieved February 3, 2018.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, *103*(5), 2670–2676.

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. *Language, Ethnicity and Intergroup Relations*, *13*, 119.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206–219.

Bradlow, A., Clopper, C., Smiljanic, R., & Walter, M. A. (2010). A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication*, *52*(11–12), 930–942. https://doi.org/10.1016/j.specom.2010.06.003

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206–219.

Britain, D. (2013). Space, diffusion and mobility. *The Handbook of Language Variation and Change*, 469–500.

Brosseau-Lapré, F., Rvachew, S., Clayards, M., & Dickson, D. (2013). Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics*, *34*(3), 419–441.

Brunellière, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language*, *125*(1), 82–93.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, *7*(4–5), 585–614.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language and Linguistics Compass*, *4*(6), 377–389.

Campbell-Kibler, K. (2011). Intersecting variables and perceived sexual orientation in men. *American Speech*, *86*(1), 52–68. https://doi.org/10.1215/00031283-1277510

Campbell-Kibler, K., & deandre, miles-hercules. (2021). Perception of gender and sexuality. In J. Baxter & J. Angouri (Eds.), *The Routledge Handbook of Language, Gender and Sexuality*, 650–666.

Carpenter, J., & Hilliard, S. (2005). Shifting parameters of individual and group variation: African American English on Roanoke Island. *Journal of English Linguistics*, *33*(2), 161–184.

Chambers, J. K. (2009). Sociolinguistic theory. Revised edition. *Malden, Mass.: Wiley-Blackwell*.

Chao, S.-C., Ochoa, D., & Daliri, A. (2019). Production variability and categorical perception of vowels are strongly linked. *Frontiers in Human Neuroscience*, *13*, 96.

Chládková, K., Podlipský, V. J., & Chionidou, A. (2017). Perceptual adaptation of vowels generalizes across the phonology and does not require local context. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 414–427. https://doi.org/10.1037/xhp0000333

Chodroff, E. (2017). *Structured variation in obstruent production and perception*. Johns Hopkins University.

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. https://doi.org/10.1016/j.wocn.2017.01.001

Chodroff, E., & Wilson, C. (2018). Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for

place and voice. *Linguistics Vanguard*, *4*(s2), 20170047. https://doi.org/10.1515/lingvan-2017-0047

Chodroff, E., & Wilson, C. (2020). Acoustic–phonetic and auditory mechanisms of adaptation in the perception of sibilant fricatives. *Attention, Perception, & Psychophysics*, *82*(4), 2027–2048. https://doi.org/10.3758/s13414-019-01894-2

Chodroff, E., & Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language*, *98*(2), 250–289.

Chodroff, E., Golden, A., & Wilson, C. (2019). Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America*, *145*(1), EL109–EL115.

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 The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.*

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change*, *7*(2), 209–228.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America*, *144*(3), EL172–EL177.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Clopper, C. G., & Pisoni, D. B. (2004a). Effects of Talker Variability on Perceptual Learning of Dialects. *Language and Speech*, *47*(3), 207–238. https://doi.org/10.1177/00238309040470030101

Clopper, C. G., & Pisoni, D. B. (2004b). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*, *16*(1), 31–48.

Clopper, C. G., & Pisoni, D. B. (2004c). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, *32*(1), 111–140. https://doi.org/10.1016/S0095-4470(03)00009-3

Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics*, *35*(3), 421–438. https://doi.org/10.1016/j.wocn.2006.06.001

Clopper, C. G., Levi, S. V., & Pisoni, D. B. (2006). Perceptual similarity of regional dialects of American English. *The Journal of the Acoustical Society of America*, *119*(1), 566–574. https://doi.org/10.1121/1.2141171

Clopper, C. G., Pisoni, D. B., & de Jong, K. J. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661. https://doi.org/10.1121/1.2000774

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, *29*(8), 1165–1175.

Cohn, A. C., & Renwick, M. E. (2021). Embracing multidimensionality in phonological analysis. *The Linguistic Review*, *38*(1), 101–139.

Commenges, D. (2015). Information Theory and Statistics: An overview. *ArXiv Preprint ArXiv:1511.00860*.

Cummings, S. N., & Theodore, R. M. (2023). Hearing is believing: Lexically guided perceptual learning is graded to reflect the quantity of evidence in speech input. *Cognition*, *235*, 105404.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). *Prelexically-driven perceptual retuning of phoneme boundaries*. *Interspeech 2008*, 2056.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*(3), 710–718. https://doi.org/10.1016/j.cognition.2008.06.003

DeCamp, D. (1953). *The pronunciation of English in San Francisco*. University of California, Berkeley.

Decker, P. D. (2010). Sounds Shifty: Gender and Age Differences in Perceptual Categorization During a Phonetic Change in Progress. *University of Pennsylvania Working Papers in Linguistics*, *15*(2), 50–60.

Docherty, G. J., & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*, *142*, 42–56.

Docherty, G. J., Langstrof, C., & Foulkes, P. (2013). Listener evaluation of sociophonetic variability: Probing constraints and capabilities. *Linguistics*, *51*(2), 355–380.

Dodsworth, R. (2018). Community Detection and the Reversal of the Southern Vowel Shift in Raleigh, North Carolina. *Language Variety in the New South: Contemporary Perspectives on Change and Variation*, 241–256.

Dodsworth, R., & Benton, R. A. (2017). Social network cohesion and the retreat from Southern vowels in Raleigh. *Language in Society*, *46*(3), 371–405.

Dodsworth, R., & Kohn, M. (2012). Urban rejection of the vernacular: The SVS undone. *Language Variation and Change*, *24*, 221–245. https://doi.org/10.1017/S0954394512000105

D'Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, *19*(2), 241–256.

D'Onofrio, A., & Benheim, J. (2020). Contextualizing reversal: Local dynamics of the Northern Cities Shift in a Chicago community. *Journal of Sociolinguistics*, *24*(4), 469–491.

D'onofrio, A., Eckert, P., Podesva, R. J., Pratt, T., & Van Hofwegen, J. (2016). The low vowels in California's central valley. *Publication of the American Dialect Society*, *101*(1), 11–32.

Dorian, N. (1994). Varieties of Variation in a Very Small Place: Social Homogeneity, Prestige Norms, and Linguistic Variation. *Language*, *70*(4), 631–696.

Drager, K. (2010). Sociophonetic Variation in Speech Perception: Sociophonetic Variation in Speech Perception. *Language and Linguistics Compass*, *4*(7), 473–480. https://doi.org/10.1111/j.1749-818X.2010.00210.x

Drouin, J. R., Theodore, R. M., & Myers, E. B. (2016). Lexically guided perceptual tuning of internal phonetic category structure. *The Journal of the Acoustical Society of America*, *140*(4), 307–313. https://doi.org/10.1121/1.4964468

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., & Martey, N. (2000-2005). Santa Barbara corpus of spoken American English. *CD-ROM. Philadelphia: Linguistic Data Consortium*.

Durian, D., & Cameron, R. (2018). Another look at the development of the Northern Cities Shift in Chicago. *NWAV (New Ways of Analyzing Variation), 47*.

Eckert, P. (1980). *Clothing and geography in a suburban high school*. Ann Arbor, Department of Anthropology, University of Michigan.

Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society*, *17*(2), 183–207.

Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. Teachers College Press.

Eckert, P. (2000). *Linguistic variation as social practice: The linguistic construction of identity in Belten High*. Wiley-Blackwell.

Eckert, P. (2008). Where do ethnolects stop? *International Journal of Bilingualism*, *12*(1–2), 25–42. https://doi.org/10.1177/13670069080120010301

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, *41*(1), 87–100. https://doi.org/10.1146/annurev-anthro-092611-145828

Eckert, P., & Labov, W. (2017). Phonetics, phonology and social meaning. *Journal of Sociolinguistics*, *21*(4), 467–496. https://doi.org/10.1111/josl.12244

Eckert, P., & McConnell-Ginet, S. (2012). Constructing meaning, constructing selves: Snapshots of language, gender, and class from Belten High. In K. Hall, & M. Bucholtz (Eds.), *Gender articulated* (pp. 469–507). Routledge.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238. https://doi.org/10.3758/BF03206487

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, *4*, 148.

Erker, D. (2017). The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language and Linguistics Compass*, *11*(1), e12232.

Erker, D., & Otheguy, R. (2016). Contact and coherence: Dialectal leveling and structural convergence in NYC Spanish. *Lingua*, *172*, 131–146. https://doi.org/10.1016/j.lingua.2015.10.011

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

Fasold, R. W. (1991). The quiet demise of variable rules. *American Speech, 66(1)*, 3–20.

Feagin, C. (1986). More evidence of major vowel change in the South. In D. Sankoff (Ed.), *Diversity and Diachrony* (pp. 83–95). John Benjamins.

Felker, E., Ernestus, M., & Broersma, M. (2019). Lexically Guided Perceptual Learning of a Vowel Shift in an Interactive L2 Listening Context. *Interspeech 2019*, 3123–3127. https://doi.org/10.21437/Interspeech.2019-1414

Fitt, S. (2000). *Documentation and user guide to UNISYN lexicon and post-lexical rules*. Tech. Rep., Centre for Speech Technology Research, Edinburgh.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research*, *38*, 379–412.

Forrest, J. (2015). Community rules and speaker behavior: Individual adherence to group constraints on (ING). *Language Variation and Change*, *27*(3), 377–406. https://doi.org/10.1017/S0954394515000137

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology*, *1*(1). https://doi.org/10.1515/labphon.2010.003

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438.

Foulkes, P., & Hay, J. B. (2015). The Emergence of Sociophonetic Structure. *The Handbook of Language Emergence*, *87*, 292.

Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics*, *55*(6), 597–610.

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., & Hagoort, P. (2017). Individual variability as a window on production-perception interactions in speech motor control. *The Journal of the Acoustical Society of America*, *142*(4), 2007–2018.

Fridland, V. (2000). The Southern Shift in Memphis, Tennessee. *Language Variation and Change*, *11*(3), 267–285. https://doi.org/10.1017/S0954394599113024

Fridland, V. (2003). "Tie, tied and tight": The expansion of /ai/ monophthongization in African-American and European-American speech in Memphis, Tennessee. *Journal of Sociolinguistics*, *7*, 279–298. https://doi.org/10.1111/1467-9481.00225

Fridland, V. (2012). Rebel Vowels: How Vowel Shift Patterns are Reshaping Speech in the Modern South. *Language and Linguistics Compass, 6(3)*, 183–192.

Fridland, V., & Bartlett, K. (2006). The social and linguistic conditioning of back vowel fronting across ethnic groups in Memphis, Tennessee. *English Language and Linguistics*, *10*(1), 1–22. https://doi.org/10.1017/S1360674305001681

Fridland, V., Bartlett, K., & Kreuz, R. (2004). Do you hear what I hear? Experimental measurement of the perceptual salience of acoustically manipulated vowel variants by Southern speakers in Memphis, TN. *Language Variation and Change*, *16*(1), 1–16. https://doi.org/10.1017/S0954394504161012

Fridland, V., & Kendall, T. (2012). Exploring the relationship between production and perception in the mid front vowels of U.S. English. *Lingua*, *122*(7), 779–793. https://doi.org/10.1016/j.lingua.2011.12.007

Fridland, V., & Kendall, T. (2015). Within-Region Diversity in the Southern Vowel Shift: Production and Perception. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.*

Fridland, V., & Kendall, T. (2018). Regional Identity and Listener Perception. In B. E. Evans, E. J. Benson, & J. Stanford (Eds.), *Language Regard* (1st ed., pp. 132–150). Cambridge University Press. https://doi.org/10.1017/9781316678381.008

Fridland, V., & Kendall, T. (2022). 18 Managing Sociophonetic Data in a Study of Regional Variation. *The Open Handbook of Linguistic Data Management*, 237.

Fridland, V., Kendall, T., & Farrington, C. (2013). The role of duration in regional US vowel shifts. *Proceedings of Meetings on Acoustics*, *19*(1).

Fridland, V., Kendall, T., & Farrington, C. (2014). Durational and spectral differences in American English vowels: Dialect variation within and across regions. *The Journal of the Acoustical Society of America*, *136*(1), 341–349. https://doi.org/10.1121/1.4883599

Fruehwald, J. (2013). *The phonological influence on phonetic change* [University of Pennsylvania]. Publicly Accessible Penn Dissertations. 862. https://repository.upenn.edu/edissertations/862

Fruehwald, J. (2017). *Gender effects on inter and intra-speaker variance in sound change*. NWAV 46, Madison, WI.

Godfrey, J., & Holliman, E. (1993). Switchboard-1 Release 2 LDC97S62. *Linguistic Data Consortium*, 34.

Gordon, M. (2005). The Midwest and West. *Handbook of Varieties of English: The Americas and Caribbean*, *1*, 338–350.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*(4), 575–590.

Grama, J., & Kennedy, R. (2019). 2. Dimensions of Variance and Contrast in the Low Back Merger and the Low-Back-Merger Shift. *Publication of the American Dialect Society*, *104*(1), 31–55.

Gumperz, J. J. (1968). *The speech community. International Encyclopedia of Social Sciences 9, 381-386. Reprinted as Ch. 7 of JJ Gumperz Language in Social Groups*. Stanford University Press.

Gumperz, J. J. (2009). The speech community. *Linguistic Anthropology: A Reader*, *1*(66), 66–73.

Gunter, K. M., Vaughn, C. R., & Kendall, T. (2020). Perceiving Southernness: Vowel categories and acoustic cues in Southernness ratings. *The Journal of the Acoustical Society of America*, *147*(1), 643–656.

Gunter, K., Vaughn, C., & Kendall, T. (2021). Contextualizing/s/retraction: Sibilant variation and change in Washington DC African American Language. *Language Variation and Change*, *33*(3), 331–357.

Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). Academic Press.

Guy, G. R., & Hinskens, F. (2016). Linguistic coherence: Systems, repertoires and speech communities. *Lingua*, *172–173*, 1–9. https://doi.org/10.1016/j.lingua.2016.01.001

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America*, *102*(1), 655–658.

Hall-Lew, L. (2009). Ethnic Practice is Local Practice: Phonetic change in San Francisco, California. *Poster Presented at Vox California*.

Hall-Lew, L. (2010). Improved representation of variance in measures of vowel merger. *The Journal of the Acoustical Society of America*, *127*(3), 2020. https://doi.org/10.1121/1.3385271

Hall-Lew, L. (2013). 'Flip-flop' and mergers-in-progress1. *English Language & Linguistics*, *17*(2), 359–390.

Harmon, Z., Idemaru, K., & Kapatsinski, V. (2019). Learning mechanisms in cue reweighting. *Cognition*, *189*, 76–88. https://doi.org/10.1016/j.cognition.2019.03.011

Herold, R. (1990). *Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania*. University of Pennsylvania.

Hay, J., & Drager, K. (2010). *Stuffed toys and speech perception*. *48*(4), 865–892. https://doi.org/doi:10.1515/ling.2010.027

Hay, J., Drager, K., & Warren, P. (2009). Careful who you talk to: An effect of experimenter identity on the production of the NEAR/SQUARE merger in New Zealand English. *Australian Journal of Linguistics*, *29*(2), 269–285.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139–2152.

Hazen, K. (2018). Listening to Rural Voices: Sociolinguistic Variation in West Virginia. In C. Mallinson & E. Seale (Eds.), *Rural voices: Language, identity, and social change across place* (pp. 75–90). Rowman & Littlefield.

Hazen, K., Lovejoy, J., Daugherty, J., Vandevender, M., Schumann, W., & Fletcher, R. (2016). Continuity and change of English consonants in Appalachia. *Appalachia Revisited: New Perspectives on Place, Tradition, and Progress*, 119–138.

Heald, S. L., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*, 35.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *95*(5), 3099–3111.

Horvath, B. M., & Horvath, R. J. (2002). The geolinguistics of/l/vocalization in Australia and New Zealand. *Journal of Sociolinguistics*, *6*(3), 319–346.

Horvath, B. M., & Horvath, R. J. (2003). A closer look at the constraint hierarchy: Order, contrast, and geographical scale. *Language Variation and Change*, *15*(02). https://doi.org/10.1017/S0954394503152015

Horvath, B., & Sankoff, D. (1987). Delimiting the Sydney speech community. *Language in Society*, *16*(2), 179–204. https://doi.org/10.1017/S0047404500012252

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939.

Idemaru, K., & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1009–1021. https://doi.org/10.1037/a0035269

Idemaru, K., & Vaughn, C. (2020). Perceptual tracking of distinct distributional regularities within a single voice. *The Journal of the Acoustical Society of America*, *148*(6), 427–432. https://doi.org/10.1121/10.0002762

Impe, L., Geeraerts, D., & Speelman, D. (2008). Mutual Intelligibility of Standard and Regional Dutch Language Varieties. *International Journal of Humanities and Arts Computing*, *2*(1–2), 101–117. https://doi.org/10.3366/E1753854809000330

Irons, T. L. (2007). On the Southern Shift in Appalachian English. *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 121–134.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson K & Mullennix J (Eds.), *Talker variability in speech processing* (pp. 145–165). Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499. https://doi.org/10.1016/j.wocn.2005.08.004

Jongman, A., & McMurray, B. (2017). On invariance: Acoustic input meets listener expectations. In A. Lahiri & S. Kotzor (Eds.), *The Speech Processing Lexicon* (pp. 21–51). De Gruyter. https://doi.org/10.1515/9783110422658-003

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252–1263.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351.

Kapatsinski, V. (2018). *Changing minds changing tools: From learning theory to language acquisition to language change*. MIT Press.

Kataoka, R., & Koo, H. (2017). Comparing malleability of phonetic category between [i] and [u]. *The Journal of the Acoustical Society of America*, *142*(1), EL42–EL48. https://doi.org/10.1121/1.4986422

Kay, M. (2020). *Ggdist:* Visualizations of distributions and uncertainty. *R package Version 3.0.0.*

Kelley, M. C., & Tucker, B. V. (2020). A comparison of four vowel overlap measures. *The Journal of the Acoustical Society of America*, *147*(1), 137–145. https://doi.org/10.1121/10.0000494

Kendall, T. (2007). Advancing the utility of the transcript: A computer-enhanced methodology. *Linguistica Atlantica*, 51–55.

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

Kendall, T., & Fridland, V. (2017). Regional relationships among the low vowels of U.S. English: Evidence from production and perception. *Language Variation and Change*, *29*(2), 245-271. https://doi.org/10.1017/S0954394517000084

Kendall, T., & Fridland, V. (2021). *Sociophonetics*. Cambridge University Press.

Kendall, T., Pharao, N., Stuart-Smith, J., & Vaughn, C. (2023). Advancements of phonetics in the 21st century: Theoretical issues in sociophonetics. *Journal of Phonetics*, *98*, 101226.

Kim, S. K., & Sumner, M. (2017). Beyond lexical meaning: The effect of emotional prosody on spoken word recognition. *The Journal of the Acoustical Society of America*, *142*(1), EL49–EL55.

King, E., & Sumner, M. (2015). Voice-specific effects in semantic association. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 6.

King, S., & Sumner, M. (2014). Voices and Variants: Effects of Voice on the Form-Based Processing of Words with Different Phonological Variants. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 36.

Kleinschmidt, D. F. (2016). *Perception in a Variable but Structured World: The Case of Speech Perception* [Preprint]. Thesis Commons. https://doi.org/10.31237/osf.io/zwves

Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience*, *34*(1), 43–68. https://doi.org/10.1080/23273798.2018.1500698

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Kleinschmidt, D. F., & Jaeger, T. F. (2016). What do you expect from an unfamiliar talker? *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 6.

Kohn, M. E., & Farrington, C. (2012). Evaluating acoustic speaker normalization algorithms: Evidence from longitudinal child data. *The Journal of the Acoustical Society of America*, *131*(3), 2237–2248.

Komsta, L., & Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis and related tests. *R Package Version 0.14.1*.

Koops, C. (2014). Iconization and the Timing of Southern Vowels: A Case Study of/{æ}. *University of Pennsylvania Working Papers in Linguistics*, *20*(2), 10.

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 12.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?. *Cognitive psychology*, *51*(2), 141-178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, *13*(2), 262–268. https://doi.org/10.3758/BF03193841

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15. https://doi.org/10.1016/j.jml.2006.07.010

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81. https://doi.org/10.1016/j.cognition.2007.07.013

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–338. https://doi.org/10.1111/j.1467-9280.2008.02090.x

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*(2), 225–248.

Kruschke, J. K. (2006). Concept learning and categorization: Models. *Encyclopedia of Cognitive Science*.

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

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

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Labov, W. (1966). *The linguistic variable as a structural unit.*

Labov, W. (1969). Contraction, Deletion, and Inherent Variability of the English Copula. *Language*, *45*(4), 715–762.

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular* (Vol. 3). University of Pennsylvania Press.

Labov, W. (Ed.). (1980). *Locating language in time and space*. Academic Press.

Labov, W. (1994). *Principles of linguistic change, Volume 1: Internal factors*. Wiley-Blackwell.

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social factors*. Wiley-Blackwell.

Labov, W. (2010). *Principles of linguistic change, Volume 3: Cognitive and cultural factors*. Wiley-Blackwell.

Labov, W., & Baranowski, M. (2006). 50 msec. *Language Variation and Change*, *18*(3), 223–240.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English phonetics, phonology, and sound change*. Walter de Gruyter.

Labov, W., Karen, M., & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, *3*(1), 33–74.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress* (Vol. 1). US Regional Survey.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98–104.

Lai, W. (2021). *The online adjustment of speaker-specific phonetic beliefs in multi-speaker speech perception*. University of Pennsylvania.

Lenth, R., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2023). Emmeans: Estimated Marginal Means, Aka Least-Squares Means. *R Package Version 1.7.0*.

Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*, 55–70. https://doi.org/10.1016/j.cognition.2018.01.003

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1783–1798. https://doi.org/10.1037/xhp0000092

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608. https://doi.org/10.1121/1.1912396

Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics*, *31*(3-4), 321-339. https://doi.org/10.1016/S0095-4470(03)00045-7

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.1894649

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(2), 391–409. https://doi.org/10.1037/0096-1523.33.2.391

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). BayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541.

Martinet, A. (1955). *Économie des changements phonétiques: Traité de phonologie diachronique.* A. Francke.

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–283). Cambridge University Press.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, *23*(4), 558–614.

Massaro, D. W., & Cohen, M. M. (1993). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication*, *13*(1–2), 127–134.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science*, *32*(3), 543–562. https://doi.org/10.1080/03640210802035357

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McAuliffe, M. (2015). *Attention and salience in lexically-guided perceptual learning* (Doctoral dissertation, University of British Columbia).

McAuliffe, M., & Babel, M. (2016). Stimulus-directed attention attenuates lexically-guided perceptual learning. *The Journal of the Acoustical Society of America*, *140*(3), 1727-1738.

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

McAuliffe, M., Coles, A., Goodale, M., Mihuc, S., Wagner, M., Stuart-Smith, J., & Sonderegger, M. (2019). ISCAN: A system for integrated phonetic analyses across speech corpora.

McLarty, J. A. (2019). *Prosodic prominence perception, regional background, ethnicity and experience: Naive perception of African American English and European American English* [Doctoral dissertation, University of Oregon].

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219–246. https://doi.org/10.1037/a0022325

McQueen, J. M., & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception (PSP2005)*, 233–236.

Ménard, L., Schwartz, J.-L., & Aubin, J. (2008). Invariance and variability in the production of the height feature in French vowels. *Speech Communication*, *50*(1), 14–28.

Mendoza-Denton, R., Goldman-Flythe, M., Pietrzak, J., Downey, G., & Aceves, M. J. (2010). Group-value ambiguity: Understanding the effects of academic feedback on minority students' self-esteem. *Social Psychological and Personality Science*, *1*(2), 127–135.

Meyerhoff, M., & Walker, J. A. (2007). The persistence of variation in individual grammars: Copula absence in urban sojourners and their stay-at-home peers, Bequia (St Vincent and the Grenadines). *Journal of Sociolinguistics*, *11*(3), 346–366. https://doi.org/10.1111/j.1467-9841.2007.00327.x

Meyerhoff, M., & Walker, J. A. (2013). An existential problem: The sociolinguistic monitor and variation in existential constructions on Bequia (St. Vincent and the Grenadines). *Language in Society*, *42*(4), 407–428. https://doi.org/10.1017/S0047404513000456

Mielke, J., Thomas, E. R., Fruehwald, J., McAuliffe, M., Sonderegger, M., Stuart-Smith, J., & Dodsworth, R. (2019). Age vectors vs. Axes of intraspeaker variation in vowel formants measured automatically from several English speech corpora. *Proceedings of the 19th International Congress of Phonetic Sciences.*

Milroy, L., & Margrain, S. (1980). Vernacular language loyalty and social network. *Language in Society*, *9*(1), 43–70.

Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation1. *Journal of Linguistics*, *21*(2), 339–384.

Milroy, L., & Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society*, *21*(1), 1–26.

Mitterer, H., & Reinisch, E. (2017). Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops. *Language, Cognition and Neuroscience*, *32*(9), 1133–1147. https://doi.org/10.1080/23273798.2017.1286361

Munson, C. (2011). *Perceptual learning in speech reveals pathways of processing* [Doctoral dissertation, University of Iowa].

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790.

Nesbitt, M. (2018). Shifting from the shift: Loss of dialect distinction in the US. *The Journal of the Acoustical Society of America*, *143*(3), 1970.

Nesbitt, M. (2021). The Rise and Fall of the Northern Cities Shift: Social and Linguistic Reorganization of TRAP in Twentieth-Century Lansing, Michigan. *American Speech*, 96(3), 332–370. https://doi.org/10.1215/00031283-8791754

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*(3), 1181–1196. https://doi.org/10.1121/1.1348009

Niedzielski, N. A. (1999). *The Effect of Social Information on the Phonetic Perception of Sociolinguistic Variables*. *18*(1), 231.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. *Talker Variability in Speech Processing*, 109–132.

Nycz, J. (2015). Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass*, *9*(11), 469–482.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017.

Oushiro, L. (2016). Social and structural constraints in lectal cohesion. *Lingua*, *172–173*, 116–130. https://doi.org/10.1016/j.lingua.2015.10.015

Oushiro, L. (2019). A computational approach for modeling the indexical field / Uma abordagem computacional para a modelagem de campos indexicais. *Revista de estudos da linguagem*, *27*(4), 1737. https://doi.org/10.17851/2237-2083.0.0.1737-1786

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309.

Peirce, J., Hirst, R., & MacAskill, M. (2022). *Building Experiments in PsychoPy*. Sage Publications.

Peterson, G. E., & Barney, H. L. (1952). *Control Methods Used in a Study of the Vowels*. *The Journal of the acoustical society of America*, *24*(2), 175–184.

Pierrehumbert, J. B. (2001). Stochastic phonology. *Glot International*, *5*(6), 195–207.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *4*(34), 516–530.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*, 33–52. https://doi.org/10.1146/annurev-linguistics-030514-125050

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release)[www.Buckeyecorpus.osu.edu]. *Ed. Columbus, OH: Department of Psychology, Ohio State University (Distributor)*.

Plichta, B., & Preston, D. (2005). The /ay/s have it: The perception of /ay/ as a north-south stereotype in United States English. *Acta Linguistica Hafniensia*, *37*, 107–130.

Plichta, B., & Rakerd, B. (2010). Perceptions of/a/-fronting across two Michigan dialects. *A Reader in Sociophonetics*, *219*, 223–240.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona 1. *Journal of Sociolinguistics*, *11*(4), 478–504.

Podesva, R. J. (2011). The California vowel shift and gay identity. *American Speech*, *86*(1), 32–51.

Podesva, R. J., D'Onofrio, A., Van Hofwegen, J., & Kim, S. K. (2015). Country ideology and the California Vowel Shift. *Language Variation and Change*, *27*(2), 157–186. https://doi.org/10.1017/S095439451500006X

Preston, D. R. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics*. De Gruyter. https://doi.org/10.1515/9783110871913

Preston, D. R. (1991). Sorting out the variables in sociolinguistic theory. *American Speech*, *66*(1), 33–56.

Preston, D. R. (1993). Variation linguistics and SLA. *Second Language Research*, *9*(2), 153–172.

Preston, D. (1996). Where the Worst English Is Spoken. In E. Schneider (Ed.), *Varieties of English Around the World: Focus on the USA* (pp. 16; 297-).

Preston, D. R. (2011). *Perceptual dialectology: Nonlinguists' views of areal linguistics* (Vol. 7). Walter de Gruyter.

Quam, C., & Creel, S. C. (2021). Impacts of acoustic-phonetic variability on perceptual development for spoken language: A review. *WIREs Cognitive Science*, *12*(5). https://doi.org/10.1002/wcs.1558

R Core Team. (2018). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. https://www.R-project.org

R Core Team. (2021). R: a language and environment for statistical computing. *Foundation for Statistical Computing*. https://www.r-project.org/Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.

Rakerd, B., & Plichta, B. (2010). More on Michigan Listeners' Perceptions of /a/-Fronting. *American Speech*, *85*(4), 431–449. https://doi.org/10.1215/00031283-2010-023

Reddy, S., & Stanford, J. N. (2015). Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard*, *1*(1), 15–28.

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

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, *45*, 91–105. https://doi.org/10.1016/j.wocn.2014.04.002

Richter, C., Feldman, N. H., Salgado, H., & Jansen, A. (2016). A framework for evaluating speech representations. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Rosenfelder, I., Fruehwald, J., Evanini K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2015). FAVE (Forced Alignment and Vowel Extraction). *Version 1.2.2*. https://doi.org/10.5281/zenodo.22281

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531. https://doi.org/10.1007/BF00973770

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Salesky, E., Chodroff, E., Pimentel, T., Wiesner, M., Cotterell, R., Black, A. W., & Eisner, J. (2020). A corpus for large-scale phonetic typology. *ArXiv Preprint ArXiv:2005.13962*.

Samuel, A. G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474. https://doi.org/10.1037/0096-3445.110.4.474

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218. https://doi.org/10.3758/APP.71.6.1207

Schwartz, J.-L., & Lucie, M. (2019). *Structured idiosyncrasies in vowel systems*.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 450–466.

Sonderegger, M., Stuart-Smith, J., Knowles, T., Macdonald, R., & Rathcke, T. (2020). Structured heterogeneity in Scottish stops over the twentieth century. *Language*, *96*(1), 94–125.

Sonderegger, M., & Stuart, J. (2022). 15 Managing Data for Integrated Speech Corpus Analysis. *The Open Handbook of Linguistic Data Management*, 195.

Stanford, J. N. (2019). *New England English: Large-scale acoustic sociophonetics and dialectology*. Oxford University Press, USA.

Stanley, J. A. (2020). *Vowel dynamics of the elsewhere shift: A sociophonetic analysis of English in Cowlitz County, Washington* [Doctoral dissertation, University of Georgia].

Staum Casasanto, L. (2010). What do Listeners Know about Sociolinguistic Variation? *University of Pennsylvania Working Papers in Linguistics*, *15*(2).

Strand, E. A. (1999). Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology*, *18*(1), 86–100. https://doi.org/10.1177/0261927X99018001006

Strand, E. A. (2000). *Gender stereotypes efffects in speech processing*. [Doctoral dissertation, The Ohio State University].

Strand, E. A., & Johnson, K. (1996). Gradient and Visual Speaker Normalization in the Perception of Fricatives. In D. Gibbon (Ed.), *Natural Language Processing and Speech Technology* (pp. 14–26). De Gruyter. https://doi.org/10.1515/9783110821895-003

Stuart-Smith, J., Sonderegger, M., Macdonald, R., Mielke, J., McAuliffe, M., & Thomas, E. (2019). Large-scale acoustic analysis of dialectal and social factors in English /s/-retraction. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1273–1277.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*(1), 131–136. https://doi.org/10.1016/j.cognition.2010.10.018

Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America*, *134*(6), EL485–EL491. https://doi.org/10.1121/1.4826151

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501. https://doi.org/10.1016/j.jml.2009.01.001

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, *4*(JAN), 1–13. https://doi.org/10.3389/fpsyg.2013.01015

Szakay, A., Babel, M., & King, J. (2016). Social categories are shared across bilinguals′ lexicons. *Journal of Phonetics*, *59*, 92–109. https://doi.org/10.1016/j.wocn.2016.09.005

Tamminga, M. (2019). Interspeaker covariation in Philadelphia vowel changes. *Language Variation and Change*, *31*(2), 119–133. https://doi.org/10.1017/S0954394519000139

Tamminga, M., & Wade, L. (2022). Coherence across social and temporal scales. In *The Coherence of Linguistic Communities*, 34–52. Routledge.

Tamminga, M., MacKenzie, L., & Embick, D. (2016). The dynamics of variation in individuals. *Linguistic Variation*, *16*(2), 300–336. https://doi.org/10.1075/lv.16.2.06tam

Tamminga, M., Wilder, R., Lai, W., & Wade, L. (2020). Perceptual learning, talker specificity, and sound change. *Papers in Historical Phonology*, *5*, 90–122. https://doi.org/10.2218/pihph.5.2020.4439

Tanner, J., Sonderegger, M., Stuart-Smith, J., & Fruehwald, J. (2020). Toward "English" Phonetics: Variability in the Pre-consonantal Voicing Effect Across English Dialects and Speakers. *Frontiers in Artificial Intelligence*, *3*, 38. https://doi.org/10.3389/frai.2020.00038

Theodore, R. M., & Miller, J. L. (2008). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, *124*(4), 2438. https://doi.org/10.1121/1.4782541

Theodore, R. M., & Miller, J. L. (2008). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, *128*(4), 2090–2099. https://doi.org/10.1121/1.3467771

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, *26*(3), 985–992. https://doi.org/10.3758/s13423-018-1551-5

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*(6), 3974–3982.

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America*, *138*(2), 1068–1078. https://doi.org/10.1121/1.4927489

Thomas, E. R. (2001). An acoustic analysis of vowel variation in New World English. *Publication of the American Dialect Society*.

Thomas, E. R. (2002). Sociophonetic applications of speech perception experiments. *American Speech*, *77*(2), 115–147.

Thomas, E. R. (2003). Secrets Revealed by Southern Vowel Shifting. *American Speech*, *78*(2), 150–170. https://doi.org/10.1215/00031283-78-2-150

Thomas, E. R. (2011). *Sociophonetics: An Introduction*. Palgrave Macmillan.

Thomas, E. R. (2019). 9. A Retrospective on the Low-Back-Merger Shift. *Publication of the American Dialect Society*, *104*(1), 180–204.

Thomas, E. R., & Kendall, T. (2007). *NORM: The vowel normalization and plotting suite.* [Online Resource].

Trudgill, P. (1986). *Dialects in Contact*. Blackwell.

Tsiplakou, S., Armostis, S., & Evripidou, D. (2016). Coherence 'in the mix'? Coherence in the face of language shift in Cypriot Greek. *Lingua*, *172–173*, 10–25. https://doi.org/10.1016/j.lingua.2015.10.014

Tzeng, C. Y., Nygaard, L. C., & Theodore, R. M. (2021). A second chance for a first impression: Sensitivity to cumulative input statistics for lexically guided perceptual learning. *Psychonomic Bulletin & Review*, *28*, 1003–1014.

Van der Zande, P., Jesse, A., & Cutler, A. (2014). Cross-speaker generalisation in two phoneme-level perceptual adaptation processes. *Journal of Phonetics*, *43*, 38–46.

Van Hofwegen, J. (2013). Vocalic style versus stylization: How outliers exemplify acoustic axes of style. *New Ways of Analyzing Variation*, *42*.

Van Hofwegen, J. (2017). *The systematicity of style: Investigating the full range of variation in everyday speech*. [Doctoral dissertation, Stanford University].

van Meel, L., Hinskens, F., & van Hout, R. (2016). Co-variation and varieties in modern Dutch ethnolects. *Lingua*, *172–173*, 72–86. https://doi.org/10.1016/j.lingua.2015.10.013

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161.

Vaughn, C. R. (2019). Expectations about the source of a speaker's accent affect accent adaptation. *The Journal of the Acoustical Society of America*, *145*(5), 3218–3232. https://doi.org/10.1121/1.5108831

Vaughn, C. R., & Kendall, T. S. (2019). Stylistically coherent variants: Cognitive representation of social meaning/Variantes estilisticamente coerentes: representação cognitiva de significados sociais. *Revista de Estudos da Linguagem*, *27*(4), 1787–1830.

Vaughn, C. R., Baese-Berk, M. M., & Idemaru, K. (2019). Re-examining phonetic variability in native and non-native speech. *Phonetica*, *76*(5), 327–358.

Villarreal, D. (2018). The Construction of Social Meaning: A Matched-Guise Investigation of the California Vowel Shift. *Journal of English Linguistics*, *46*(1), 52–78. https://doi.org/10.1177/0075424217753520

Wade, L. (2017). The role of duration in the perception of vowel merger. *Laboratory Phonology*, *8*(1), 30.

Wade, L. (2022). Experimental evidence for expectation-driven linguistic convergence. *Language*, *98*(1), 63–97. https://doi.org/10.5334/labphon.54

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of Acoustic Variability in the Perceptual Learning of Non-Native-Accented Speech Sounds. *Phonetica*, *64*(2–3), 122–144. https://doi.org/10.1159/000107913

Walker, A., & Hay, J. (2011). Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology*, *2*(1). https://doi.org/10.1515/labphon.2011.007

Walker, J. A., & Meyerhoff, M. (2013). *Studies of the Community and the Individual* (R. Bayley, R. Cameron, & C. Lucas, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199744084.013.0009

Warren, P., & Hay, J. (2006). Using sound change to explore the mental lexicon. *Cognition and Language: Perspectives from New Zealand*, *105*.

Warren, P., Hay, J., & Thomas, B. (2007). The loci of sound change effects in recognition and perception. *Laboratory Phonology*, *9*, 87–112.

Watt, D. J. (2000). Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change*, *12*(1), 69–101.

Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation*. [Doctoral dissertation, The Ohio State University].

Weatherholtz, K., & Jaeger, T. F. (2016). Speech Perception and Generalization Across Talkers and Accents. In K. Weatherholtz & T. F. Jaeger, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.95

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical foundations for a theory of language change*. University of Texas Press.

Wells, J. C. (1982). *Accents of English: Volume 1*. Cambridge University Press.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291–298.

Winn, M. (2014). Make formant continuum [Praat script]. *Version August*.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556. https://doi.org/10.3758/s13414-012-0404-y

Wolfram, W. A. (1969). Linguistic correlates of social differences in the Negro community. In J. E. Alatis (Ed.), *Report of the Twentieth Annual Roundtable Meeting on Linguistics and Language Studies: Linguistics and the Teaching of Standard English to Speakers of Other Languages and Dialects.* (pp. 249–258). Georgetown University Press.

Wolfram, W., & Beckett, D. (2000). The role of the individual and group in earlier African American English. *American Speech*, *75*(1), 3–33. https://doi.org/10.1215/00031283-75-1-3

Wright, B. A., Baese-Berk, M. M., Marrone, N., & Bradlow, A. R. (2015). Enhancing speech learning by combining task practice with periods of stimulus exposure without practice. *The Journal of the Acoustical Society of America*, *138*(2), 928–937.

Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, *97*, 30–46. https://doi.org/10.1016/j.jml.2017.07.005

Xie, X., Earle, F. S., & Myers, E. B. (2018). Sleep facilitates generalisation of accent adaptation to a new talker. *Language, Cognition and Neuroscience*, *33*(2), 196–210. https://doi.org/10.1080/23273798.2017.1369551

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PloS One*, *8*(9), e74746. https://doi.org/10.1371/journal.pone.0074746

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*, *123*(5), 3878.

Zhang, X., Wu, Y. C., & Holt, L. L. (2021). The Learning Signal in Perceptual Tuning of Speech: Bottom Up Versus Top-Down Information. *Cognitive Science*, *45*(3). https://doi.org/10.1111/cogs.12947

Zhao, Y. (2010). *Statistical inference in te learning of novel phonetic categories*. [Doctoral dissertation, Stanford University].