Constraints on Articulatory Variability: Audiovisual Perception of Lip Rounding

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Jonathan Eric Havenhill, M.S.

Washington, DC
July 30, 2018

Constraints on Articulatory Variability:
Audiovisual Perception of Lip Rounding

Jonathan Eric Havenhill, M.S.

Dissertation Advisor: Elizabeth C. Zsiga, Ph.D.

Abstract

What are the factors that shape linguistic sound systems? Perceptibility of the acoustic signal has long been argued to play a role in phonological organization. Some theories of historical change (Ohala 1981, 1993; Blevins 2004) argue that sound change results from misperception of the acoustic signal, while teleological models of phonology (Lindblom 1990; Hayes, Kirchner, and Steriade 2004) posit that speech is optimized (in part) for auditory perceptibility. Nevertheless, it is well known that speech perception is influenced by a range of non-auditory cues (McGurk and MacDonald 1976; Gick and Derrick 2009; Mayer et al. 2013).

This dissertation investigates the role of audiovisual perception in constraining patterns of articulatory variation, in which speakers employ differing articulatory strategies to achieve the same acoustic output. Such variation is widely documented for sounds like English /ɹ/ (Delattre and Freeman 1968), but does not arise in some cases where it is hypothetically possible (Harrington, Kleber, and Reubold 2011). Three experiments test the hypothesis that the range of possible variation is constrained by the availability of visual speech cues.

The first experiment considers the case of back vowel fronting in American English, where articulatory variation is predicted to be possible but has not been observed in some other varieties. It is shown that, for speakers from Southern California and South Carolina, /u/ and /o/ have retained their lip rounding as they have undergone fronting. The second experiment focuses on the fronting of /ɑ/ and /ɔ/ associated with the Northern Cities Shift.

It is found that Chicago speakers vary in the extent to which these vowels differ acoustically, but that most speakers retain the labial distinction between /ɑ/ and /ɔ/, even if the lingual distinction is lost. The third experiment demonstrates that the visual lip rounding cue enhances perception of the COT-CAUGHT contrast, making visibly round variants of /ɔ/ perceptually more robust than unround variants. It is argued that speakers prefer articulatory strategies that are contrastive in both the auditory and visual domains. This preference has typological consequences for phonological systems, such that labial segments tend to retain their labiality in diachronic change.

INDEX WORDS: phonetics, ultrasound tongue imaging, articulatory variation, audiovisual speech perception, sound change

Acknowledgments

Like any dissertation, this one could not have been completed if not for the contributions of countless others. I am extremely grateful to my advisor, Lisa Zsiga, whom I can't thank enough for all the support and opportunities she's given me over the past six years. When I joined the PhD program at Georgetown, I had only just started my studies in linguistics, but she nevertheless took me on as a student and I have since learned an enormous amount from working with her. I hope I can one day do the same for my students.

Youngah Do influenced my learning and interests in linguistics as much as anyone, and I'm glad to have had the chance to attend all of the thought-provoking seminars she led during her time at Georgetown. Working with her has been especially fulfilling, and I look forward to many more years of collaboration and friendship in Hong Kong.

This project got its start in Jen Nycz's Sociophonetics course, and would have gone nowhere if not for her enthusiastic support beginning with my final project proposal in that class. Her thoughtful comments and sociolinguistic perspective have helped at every step of the way, and I am grateful to her for introducing me to R, which made possible the data analyses and visualizations presented here.

Aside from my committee members, my thinking and learning were shaped by many of the linguistics faculty at Georgetown. I would especially like to thank David Lightfoot, Héctor Campos, and Ruth Kramer, whose support and encouragement have been central to my academic growth and made it possible to pursue a career in linguistics. I must also thank the professors and advisors at Grand Valley whose encouragement led me to study

linguistics in the first place, especially Kathryn Remlinger, Donovan Anderson, Jim Scott, Regina Smith, and Shinian Wu.

The experiments presented in this dissertation took me to several phonetics labs across the country, and it would not have been possible to carry out these experiments if not for the generous help of others. I am grateful to everyone who offered their company, advice, lab space, and help with participant recruitment. Many, many thanks to Pam Beddor, Will Styler, Eric Holt, Sharon Rose, Marc Garellek, Eric Baković, Jennifer Cole, Matt Goldrick, Annette D'Onofrio, and Chun Chan.

I could not have completed this dissertation without the support of my friends, especially Ryan Iseppi, Chelsea Champlin, Erik Greene, Chris Remijan, Florian Rott, and Amy Bush—thank you for all your support and fun times over the years/decades. Colleen Diamond and Grace Sullivan Buker were my first friends in DC, and I'm very grateful that we've remained good friends ever since. I've made far too many friends during my time at Georgetown to name them all here, but special thanks go to Shannon Mooney, Stacy Petersen, and all the members of PhonLab.

To Mom, Dad, and Kyle, thank you for your endless support, regardless of where my interests have taken me. To Tina, thank you for keeping me company during the many late nights I spent working on this dissertation, even if you spent most of that time trying to sit on my keyboard. And finally, to Liz, thank you for putting up with me over the past few years and for joining me on this roller coaster ride. I could never thank you enough for all you've done, and I can't wait to see where this next adventure takes us.

For

Arthur J. Shivers, Jr.

and

John R. Havenhill

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

> Visually perceived articulations (for example, lip movements) leave a stronger and more permanent impression than invisible articulations produced inside the vocal apparatus.
>
> — Baudouin de Courtenay (1895, 265)

This dissertation is an investigation of the articulatory and perceptual factors that constrain patterns of articulatory variation and sound change. Speech perception has long played a central role in phonological theory, as well as in theories of diachronic sound change. However, approaches in both domains typically focus on auditory speech perception and on the acoustic properties of speech more generally. For instance, Ohala (1993) and Blevins (2004) argue that sound change results from misperception of the acoustic signal, while teleological models of phonology (Lindblom 1990; Hayes, Kirchner, and Steriade 2004) posit that speech is optimized, in part, for auditory perceptibility. This focus is generally well justified, given that sound is the primary (and often only) means by which spoken language is transmitted. Nevertheless, it has long been known that listeners are sensitive to cues from a wide variety of perceptual modalities, including visual, haptic, vibrotactile, aerotactile, and proprioceptive feedback, among others. The argument presented here is that non-auditory speech perception, and visual speech perception in particular, contributes to the organization of phonological systems in at least two respects. First, visual cues can constrain patterns of articulatory variation, in which speakers employ differing articulatory strategies to achieve the same acoustic output. Second, by providing language learners with unambiguous input

with respect to a sound's place of articulation, visual cues can inhibit misperception-based sound change, leading labial segments to retain their labiality in diachronic change.

It is widely recognized that diachronic sound change originates in synchronic phonetic variation (see, e.g., Ohala 1989; Labov 1963; Labov, Yaeger, and Steiner 1972; Pierrehumbert 2001, among many others). Among the many ways in which sounds vary is in their articulation. This sort of variation can arise when two or more articulatory configurations have similar effects on the acoustic output. In order to generate an acoustic output, speakers may choose from among these articulatory configurations, either categorically or determined by speaker-specific rules (Mielke, Baker, and Archangeli 2010). Several examples of articulatory variation are known in the literature, and this sort of variation has been argued to be a driving force in sound change (Baker, Archangeli, and Mielke 2011; McGuire and Babel 2012). However, the mechanisms governing articulatory variation are not entirely understood. Assuming that two articulatory variants are acoustically equivalent, what factors contribute to a speaker favoring one articulatory configuration over another? Some researchers have argued that physiological factors, such as palate shape (Brunner, Fuchs, and Perrier 2009; Bakst and Lin 2015), or differences in perceptual acuity (Gluth and Hoole 2015) may play a role in determining the articulation of a given sound, and both factors are certainly plausible. Here it is argued that visual speech cues can also restrict the range of possible articulatory variation, both by making the speaker's articulation clear to listeners and language learners, as well as by enhancing perceptual contrasts between acoustically similar sounds, preserving weak acoustic contrasts which might otherwise be lost to sound change.

Support for this argument comes from three experiments. The first is an articulatory study of back vowel fronting in two varieties of American English. The fronting of the vowels /u/ and /o/ is observed in numerous varieties of English, but this change has been analyzed mainly in terms of acoustics rather than articulation. Work by Harrington, Kleber, and Reubold (2011) has shown that speakers of Southern Standard British English (SSBE)

achieve /u/-fronting entirely through tongue repositioning, rather than a reconfiguration of the lips, even though both articulations can produce an increase in F2 by shortening the front cavity of the vocal tract. While the fronting of back vowels in American English has been the object of much sociolinguistic inquiry, the articulatory strategies used to produce these sound patterns have not been systematically investigated. The experiment presented here provides evidence that speakers from both Southern California and South Carolina, two regions where the fronting of /u/ is particularly advanced, retain the lip rounding gesture for fronted /u/ and /o/. Thus, as for speakers of SSBE, the increase in F2 associated with vowel fronting is the result of tongue fronting, rather than unrounding of the lips.

The second experiment is an investigation of the production of fronted /ɔ/ in Chicago. In the Great Lakes region of the United States, the Northern Cities Shift has resulted in the fronting of /ɑ/ and /ɔ/, such that they, like fronted /u/, exhibit a raised F2. As with the fronting of the non-low back vowels, an increase in the F2 of /ɔ/ may be achieved either by fronting the tongue or unrounding the lips, and previous research has observed that speakers from Metro Detroit vary in their articulatory strategies (Havenhill and Do 2018). The experiment presented here seeks to determine whether speakers from Chicago exhibit the same sort of articulatory variation seen in Michigan, or whether the strategy of lip unrounding is dispreferred due to the loss of visual speech cues. It is found that all but one of the speakers who maintain an acoustic contrast between /ɑ/ and /ɔ/ do so by producing an articulatory distinction between the two vowels in terms of lip rounding. Moreover, in careful speech, the majority of speakers are shown to enhance the lip rounding distinction between /ɑ/ and /ɔ/, either by increasing the degree of lip rounding for /ɔ/, or by further unrounding /ɑ/ through an increase in lip spread. Thus, lip rounding is shown to be crucial to maintaining the COT-CAUGHT contrast among Chicagoans.

The third experiment specifically examines the role of visual lip rounding cues in maintaining the COT-CAUGHT contrast and in restricting the range of articulations for /ɔ/. This

experiment tests whether visible lip rounding on /ɔ/ improves perceivers' ability to distinguish /ɔ/ from /ɑ/. It is found that, for the majority of speakers who produce /ɔ/ with round lips in their own speech, a loss of visible lip rounding on /ɔ/ makes this vowel significantly more likely to be perceived as /ɑ/. It is therefore argued that speakers avoid the articulatory strategy of unrounding /ɔ/, because it would result in a loss of perceptual contrast with /ɑ/. In combination, these experiments suggest that, given a choice between two acoustically equivalent articulatory strategies, speakers will choose the articulation that optimizes perceptibility in both the auditory and visual domains.

It is furthermore argued that visual speech cues shape phonological typology through the mechanisms of diachronic sound change. Numerous sound changes have been attributed to the misperception of acoustically ambiguous speech sounds (see, e.g., Ohala 1981; 1993; Beddor, Krakow, and Goldstein 1986; Guion 1998; Blevins 2004, among many others). Yet, despite the well-known influence of visual cues on speech perception, few studies have considered whether visual perception can influence the direction of misperception-based sound change. It is argued here that, for sound changes involving labial consonants, visual speech cues can inhibit misperception by providing language learners with a clear cue to the segment's place of articulation. For instance, in the case of labials with secondary palatalization, the presence of a visible lip closure belies the high F2 observed in the stop release, which signals a coronal place of articulation. The visibility of the lip closure provides perceivers with a clear indication that the primary place of articulation is labial, not coronal, making them less likely to misperceive /pʲ/ as /t͡ʃ/, as compared to /tʲ/ > /t͡ʃ/ and /kʲ/ > /t͡ʃ/. It is argued that visual speech perception can account for findings from typological research (Bateman 2011; Kochetov 2011) which show that full palatalization of labials is exceptionally rare, while full palatalization of coronals and dorsals is common.

The dissertation is structured as follows. The remainder of this chapter serves as a review of relevant work in sound change, speech perception, and articulatory variation.

Chapter 2 discusses the fronting of /u/ and /o/ in English, and presents the results of a study investigating the articulation of fronted back vowels in two varieties of American English. Chapters 3 and 4 consider the relationship between articulatory variation, sound change, and audiovisual speech perception through two experiments investigating the production (Chapter 3) and perception (Chapter 4) of the Northern Cities Shift among speakers from Chicago. Chapter 5 reviews historical sound changes involving labial consonants, and discusses the role of visual speech cues in influencing the direction of these changes. Chapter 6 considers the implications of audiovisual speech perception for theories of sound change, and offers concluding remarks and directions for future research.

1.1 Speech Perception in Phonological Theory

Misperception of speech sounds has been recognized as an important source of sound change at least since the nineteenth century (Paul 1890; Baudouin de Courtenay 1895). Baudouin de Courtenay (1895), for instance, writes of "the importance of errors in hearing (*lapsus auris*), when one word is mistaken for another, as a factor of change at any given moment of linguistic intercourse and in the history of language as a social phenomenon" (267). His call for the use of experimental methods in uncovering the "types and directions" of misperception was taken up by Ohala (1981), and has since developed into one of the predominant means of investigating sound change in modern phonology. Moreover, the last several decades have seen speech perception take a central role in phonological explanation, owing to advances both in laboratory techniques and theoretical formalisms. Drawing on the tradition of Baudouin de Courtenay and Ohala, Blevins (2004) argues for a theory of phonology in which sound change is understood in terms of its similarity to processes of biological evolution, where accidental misperception is akin to random genetic mutation. In contrast, Lindblom (1990) argues that speakers are attuned to fine-grained acoustic and

articulatory detail, and actively optimize speech patterns based on the perceptual needs of the listener and on the speaker's own articulatory needs. Work has continued along these lines under the framework of Phonetically Based Phonology (Hayes, Kirchner, and Steriade 2004), in which constraints governing perceptual and articulatory demands, such as MINDIST and LAZY, are encoded in an Optimality theoretic grammar.

The approaches of Ohala (1981, 1993) and Blevins (2004) on the one hand, and Lindblom (1990) and Hayes, Kirchner, and Steriade (2004) on the other, differ in several respects. First, these approaches differ in the respective roles of the speaker and the listener. Ohala's model of innocent misapprehension and Blevins's Evolutionary Phonology can be considered listener-based, because they posit that the primary mechanism of sound change lies in the behavior of the listener. That is, whether or not a sound change takes places is dependent on whether the listener accurately perceives and parses the speaker's intended message. Meanwhile, Lindblom's theory of hyper- and hypo-articulation (H&H theory) and Phonetically Based Phonology are speaker-based, in that sound change is predicted to arise as the result of the speaker's attempt to optimize their speech based on articulatory and perceptual demands. When confronted with a speech pattern that is non-optimal in one of these respects, the speaker or language learner may reanalyze that pattern, substituting it with one that is easier to perceive or easier to produce. The goals of the speaker and listener also distinguish these two approaches. Speaker-based approaches emphasize that speakers are goal-driven and play an active role in optimizing the sound patterns of their language; as such, these approaches can be characterized as optimizing or teleological. Listener-based frameworks, on the other hand, are non-optimizing and non-teleological, because listener-based misperception is inadvertent rather than intended. In fact, Ohala (1993) argues that sound change runs contrary to "the whole purpose of the listener's interpretive activity," which is "to *preserve*, *not to change*, the pronunciation norm" (262, emphasis original). While optimal sound patterns can (and do) arise under a non-teleological approach, this

is not because optimization is the goal, but because optimal sound patterns are simply more likely to be accurately perceived by the listener. This section presents a discussion of the role of speech perception in phonetic and phonological theory as it relates to sound change, including Ohala's model of innocent misapprehension, Evolutionary Phonology, H&H theory, and Phonetically Based Phonology.

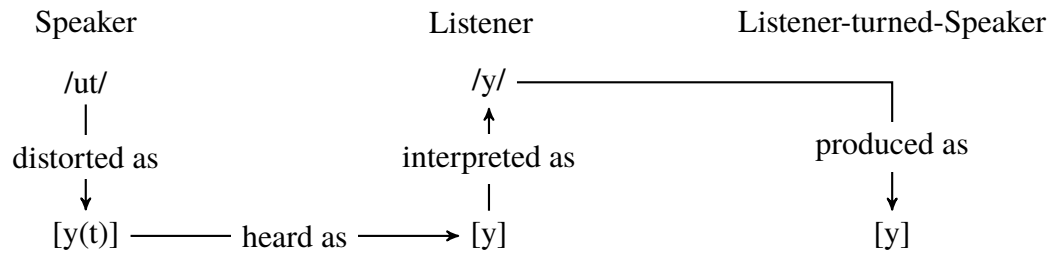
1.1.1 Innocent Misapprehension

In his listener-based theory of sound change, Ohala (1981, 1989, 1993) argues that phonetic change arises in part due to "innocent misapprehension" of an inherently ambiguous acoustic signal. He notes that due to the many-to-one mapping from articulation to the acoustic signal, listeners (or learners) may attribute a particular sound to a different articulation than that used by the speaker. In addition to the issue of articulatory mapping, coarticulatory effects and other perturbations beyond the control of the speaker, as well as channel noise, present a challenge to the listener in determining a speaker's intended utterance. Ohala suggests that because much of this variation is predictable, it can under normal circumstances be corrected for by the listener, based on their own experience with speaking the language. Evidence for such correction comes from perceptual experiments in which listeners are less likely to identify a vowel on a /i/~/u/ continuum as fronted if it occurs in a coronal context. When the same vowel appears in a labial context, it is more likely to be identified as /i/ than as /u/. The listener can attribute a raised F2 to coarticulatory effects when a fronted vowel appears in a coronal context, but cannot do so when the vowel appears in a labial environment, which has a lowering effect on F2.

Due to such coarticulatory effects, /ut/ is often realized with a pronunciation more like [yt] in natural speech. Because listeners are able to correct for this variation, the [yt] signal is typically perceived as /ut/. In some cases, such as if the conditioning environment is lost due to lenition or if the listener has had insufficient exposure to the language (as in the case of

child learners), the listener may be unable to reconstruct the speaker's intended utterance. Such failure to compensate, or hypocorrection, results in the listener acquiring /y/ as the underlying form. Subsequently, the listener will produce the vowel with /y/ as the target, and the seed for sound change will have been planted. Ohala (1981, 183) schematizes this scenario with the diagram presented in (1):

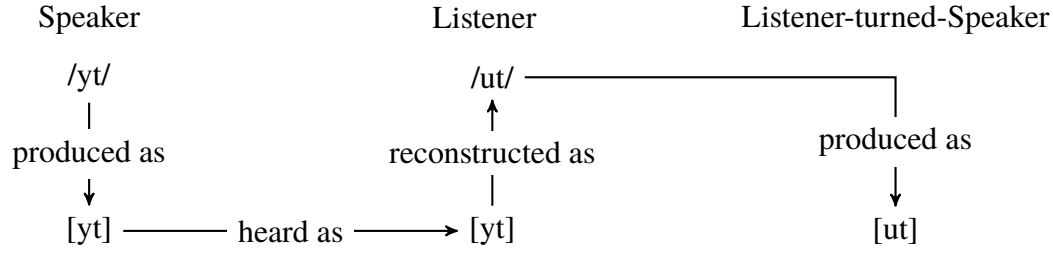
(1)



Many common types of sound change have been argued to be instances of hypocorrection, such as the influence of nasalization on the perception of vowel height, as investigated by Beddor, Krakow, and Goldstein (1986) and Krakow et al. (1988). Beddor and colleagues (1986) note that in French and many other languages, nasal vowels exhibit a difference in height relative to the oral vowels with which they alternate. This effect is argued to result from the acoustic interaction of the first formant with the spectral peak introduced by nasalization, which results in a raising or lowering of the spectral center of gravity relative to the oral vowel. In a perceptual experiment, Krakow et al. (1988) find that American English listeners are able to correctly perceive the height of both oral and nasal vowels, but the latter only when the proper conditioning environment for nasalization (i.e., a following nasal stop) is present. This finding supports the claim that listeners correct for coarticulatory effects under normal circumstances, when the acoustic signal matches the listener's expectation. However, they find that when the conditioning environment is lost, listeners fail to correct for the effects of nasalization and misperceive the vowel's height. This finding supports Ohala's (1981) claim that sound change can occur as a result of listeners' failure to reconstruct the speaker's intended utterance when the conditioning environment is lost.

In addition to cases of hypocorrection, listeners may also hypercorrect, by mistakenly attributing to coarticulatory effects some part of the acoustic signal which was intended by the speaker. Ohala (1981, 187) schematizes this scenario, the reverse of the hypocorrection, as given in (2):

(2)



Despite the fact that the speaker's intended /yt/ is correctly perceived as [yt], the listener incorrectly assumes that the source of the frontedness of the vowel is the following coronal consonant, and corrects for this. The listener acquires /ut/ as the underlying form and subsequently produces it in this manner.

As an example of hypercorrection, Ohala (1993) presents the change from Latin *quinque* /kʷiŋkʷeː/ 'five' to Italian *cinque* /t͡ʃiŋkʷe/. Ohala suggests that the loss of lip-rounding on the initial /kʷ/ was the result of confusion on the part of the listener as to the source of rounding. If the listener mistakenly interpreted the rounding on initial /kʷ/ as non-distinctive coarticulation in anticipation of the second /kʷ/, they would posit an underlying /k/ as the onset of the first syllable. Ohala argues that many such dissimilatory changes can be accounted for under the rubric of hypercorrection, but that hypercorrection (and thereby dissimilation) can occur only when features with extended temporal domains are involved. As well as rounding and labialization, such phonetic features include pharyngealization, retroflexion, glottalization, laterality, and place of articulation, among others. However, as manner of articulation is not known to extend to neighboring segments, Ohala predicts that features such as frication and continuancy will not trigger dissimilation.

1.1.2 Evolutionary Phonology

Building on the work of Ohala, Blevins (2004, 2006) proposes a theory of Evolutionary Phonology (EP), in which regular sound change is viewed in terms of its similarity to processes observed in biological evolution, such as parallel evolution and direct inheritance, among others. Just as biological traits may independently evolve in related species in response to common environmental pressures, common sound patterns emerge independently in diverse languages as a result of the universal properties of speech production and perception. Because language acquisition occurs across a noisy channel, sound changes may arise as a result of imperfect transmission of the language from one generation to the next, which Blevins suggests can be compared to mutations that occur due to errors in the replication of DNA. Certain sounds are inherently more confusable than others and are thus more likely to undergo change, leading to similar sound patterns across the world's languages. Blevins therefore argues that synchronic patterns are the result of diachronic change (or evolution), and that common typological patterns arise as the result of evolutionary convergence, grounded in universal principles of speech perception and confusability.

Like Ohala, Blevins argues that the majority of sound changes are attributable to misinterpretation or misperception on the part of the listener, and that these changes are phonetically motivated. She proposes a three-way typology of sound changes, as given in (3):

(3) General typology of sound change in Evolutionary Phonology (Blevins 2004, 32):

   i. CHANGE: The phonetic signal is misheard by the listener due to perceptual similarities of the actual utterance with the perceived utterance.

   ii. CHANCE: The phonetic signal is accurately perceived by the listener but is intrinsically phonologically ambiguous, and the listener associates a phonological form with the utterance which differs from the phonological form in the speaker's grammar.

iii. CHOICE: Multiple phonetic signals representing variants of a single phonological form are accurately perceived by the listener, and due to this variation, the listener (a) acquires a prototype or best exemplar which differs from that of the speaker; and/or (b) associates a phonological form with the set of variants which differs from the phonological form in the speaker's grammar.

Changes classified as CHANGE are largely equivalent to those captured by Ohala's hypocorrection. CHANGE occurs when the listener misperceives the speaker's intended utterance, and subsequently acquires a different pronunciation than that of the speaker. Blevins points out that in contrast to other types of sound change, CHANGE always involves a change in pronunciation from speaker to listener, but does not necessarily require a change in the underlying phonological representation. If there is evidence for the sound uttered by the speaker elsewhere in the language, the listener may instead posit a rule of alternation which modifies the sound in the environment in which it was misperceived. CHANCE, which is similar to Ohala's hypercorrection, encompasses changes in which the listener correctly perceives the speaker's intended utterance, but arrives at a different phonological representation than that of the speaker. Like hypercorrection, CHANCE typically involves features with extended temporal domains, providing the listener with the opportunity to assign some perceived phonetic characteristic to a different underlying sound than that of the speaker. Finally, CHOICE comprises changes that are associated with intraspeaker variation. Blevins notes that both CHANCE and CHANGE require idealized models of speaker-listener interactions. However, CHOICE recognizes that listeners encounter a great deal of language variation during acquisition. Rather than changing the pronunciation of a sound or its underlying representation, the listener may simply choose a different variant as the prototype form than the speaker, but still acquire a range of variants.

Blevins specifically argues against teleological approaches in part on the basis of the existence of non-optimizing sound changes. She observes for instance that while Slavic underwent a metathesizing change of the type VR > RV (where R is a liquid and V is a vowel), the Romance language Le Havre exhibits the opposite change, RV > VR, and neither change is clearly better than the other. Blevins and Garrett (2004) further argue that in order to uphold an optimizing model of sound change, one must either prove the existence of sound changes that are optimizing but not attributable to misperception, or demonstrate that a misperception-based model of change cannot account for linguistic typology. As in biological evolution, where genetic mutations do not necessarily result in adaptation, sound changes do not necessarily result in improvement over the previous system. Nevertheless, certain apparently optimal sound patterns are frequent among the world's languages, while other, less optimal patterns are rare, a fact which Blevins considers in terms of Darwinian evolution. In biology, random mutations are filtered by the process of natural selection; those favoring survival increase the likelihood that the mutation will be spread via reproduction, while those disfavoring survival are eliminated. Similarly, sounds that are easy to correctly perceive or to produce tend to be retained, while sounds that are difficult to produce and perceive are more likely to be replaced by other sounds.

1.1.3 H&H Theory

Whereas Evolutionary Phonology takes the position that sound change (and phonology itself) is non-teleological, Lindblom and colleagues (1990; 1995) have argued for optimizing theories of sound change. Under the theory of Hyper- and Hypo-articulation (H&H theory), intraspeaker variation is argued to be under the control of the speaker, and is the result of system-oriented and output-oriented constraints on speech production and perception. Lindblom argues that speech exists along a continuum between hyper- and hypo-articulate speech. On the hypo-articulate end of the spectrum, speakers consciously choose economic

gestures in order to conserve energy. Lindblom argues that this is a general property of motor behavior, in which the motor system defaults to movements which minimize displacement. On the hyper-articulate end of the spectrum, speakers are concerned with achieving the communicative goals of speech, i.e., transmitting a perceptible message to an interlocutor. Hyper- and hypo-articulate forms of speech are in competition, because while hyperarticulate speech improves perceptibility, it comes with increased articulatory cost.

Lindblom (1990) argues that this interplay can be observed in a number of aspects of speech production. First, he provides an account of coarticulation under this view, arguing that coarticulation represents a low-cost form of hypospeech, the purpose of which is to ease production. However, speakers are able to produce clear speech when necessary for communicative purposes. Clear speech differs from casual speech not only in increased volume and length, but also in changes in formant patterns such that intelligibility may be improved. Finally, he suggests that H&H theory can explain typological patterns, particularly in understanding how segment inventories are distributed. He notes that languages with smaller inventories employ only 'basic' segments, and that segments with more complex articulations are found only in larger inventories. Because complex articulations invoke a greater articulatory cost, they are found only when perceptual considerations demand them.

1.1.4 Phonetically Based Phonology

Like Lindblom's H&H theory, the framework of Phonetically Based Phonology (PBP; Hayes 1997; Hayes, Kirchner, and Steriade 2004) can be characterized as a teleological approach to phonological explanation. Under PBP, phonological markedness is argued to be the result of shared knowledge among speakers about the aspects of articulatory and perceptual processes that impede or enhance communication. Because such factors are universal and experienced by all speakers of all languages, each speaker independently posits similar constraints when acquiring their language. While not strictly a theory of

sound change, PBP makes a number of predictions with regard to how sound systems are organized and what types of sound changes are expected to occur.

The predictions made by PBP are concerned primarily with the types of changes referred to as hypercorrection (Ohala 1993) or as CHOICE/CHANCE (Blevins 2004). This is because in the case of hypocorrection or CHANGE, the listener fails to correctly perceive the speech signal at a fundamental level. The ease of articulation or relative perceptibility of the sound actually produced by the speaker is irrelevant to the listener, because they did not perceive that sound to begin with. In the case of hypercorrective changes, however, the listener is presented with an opportunity. Having correctly perceived the speech signal uttered by the speaker, the listener is now faced with parsing an inherently ambiguous signal and selecting an appropriate phonological and articulatory mapping. One way in which a listener may modify the language is by improving an alternation that is suboptimal in some way, a possibility which Steriade (2001) discusses in relation to her proposal for the P-map, a map of the speaker's knowledge of the relative perceptibility of various input-output pairs. She argues that speakers are "not averse to linguistic innovation, insofar as it remains covert" (13), proposing that speakers can make small, optimizing changes to their production, as long as their speech remains perceptible to the listener.

1.2 Multimodal Speech Perception

While each of the above theories takes perceptual knowledge to be a central component of listeners' phonetic and phonological competence, the role of perception in most linguistic research is limited to the auditory domain. A wealth of evidence suggests, however, that non-auditory speech cues, including visual, somatosensory, and tactile cues, can significantly impact listeners' perception of the speech signal. Among these, visual speech perception has received the greatest attention in the literature and is the source of the well-known McGurk

effect (McGurk and MacDonald 1976), in which mismatched auditory and visual speech cues are perceived as a fusion of the two input signals. The McGurk effect and other findings in multimodal speech perception have been an important source of evidence and debate in theories of speech perception. The following section describes the role of visual and other sensory cues in speech perception, and considers the interpretation of these findings under various theories of speech perception.

1.2.1 Visual Perception

The integration of visual cues in speech perception has been documented at least since the 1950s, when Sumby and Pollack (1954) demonstrated that the availability of visual cues improves speech intelligibility under noisy conditions. Participants in this experiment were seated in front of a speaker who read a list of bisyllabic words, which participants were asked to identify. Participants wore a headset playing white noise at one of six levels of intensity between -30 and 0 dB signal-to-noise ratio (SNR). Half of the participants faced the speaker, such that they could view the speaker's facial movements, while half faced away from the speaker. When the SNR was high, participants in both audiovisual and auditory conditions correctly identified nearly 100% of the presented words. However, these conditions diverge as the amount of noise increases. In the audiovisual condition, identification was between 40% and 100% correct with a low SNR, depending on the number of words the participants were asked to identify. In contrast, correct identification in the auditory-only condition was less than 20% with a low SNR, regardless of vocabulary size. Although visual cues provided no improvement to intelligibility when the acoustic signal was clear, Sumby and Pollack note that communication in noisy conditions is the usual case, and it is here where visual cues provide the greatest contribution to speech perception.

The McGurk effect, first reported by McGurk and MacDonald (1976), is an especially important demonstration of the effects of visual cues on speech perception. McGurk and

MacDonald observed that when adults were presented with audio recordings of the syllable [ba] paired with video of the lip movements for [ga], the syllable was perceived as [da], which they describe as a perceptual fusion. These observations were confirmed experimentally by presenting listeners with audio stimuli paired with incongruous video, such that the place of articulation heard auditorily differed from that seen visually. The auditory stimuli contained the syllables [ba], [ga], [pa], and [ka], which were respectively paired with video for [ga], [ba], [ka], and [pa]. In a second condition, participants were presented with auditory stimuli alone, without paired video. McGurk and MacDonald found the fusion effect in the incongruous audiovisual condition to be quite robust—98% of adults and 81% of three- to five-year-old children responded with [da] to the [ba]-audio/[ga]-video pairing. Interestingly, however, fusion did not occur when the video contained [ba] or [pa]. In these cases, the majority of participants perceived either a combination of these sounds, [bga], or a percept resembling the visual cue, [ba], demonstrating that listeners are attuned to visual labial cues when they are present. In light of these results, McGurk and MacDonald argue that auditory-only models of speech perception are inadequate, and that visual effects must be taken into account.

Noting that most work on the McGurk effect concerns the perception of consonants, Traunmüller and Öhrström (2007a) present an experiment investigating whether perceptual fusions occur with vowels as well. Specifically, they consider how information from both auditory and visual channels contributes to perception of vowel height and roundedness. Listeners are found to perceive an acoustically unround vowel as round when it is paired with a rounded visual stimulus, and they suggest that the acoustic signal alone is not sufficient for distinguishing /y/ from /i/ in Swedish. Native Swedish speakers were presented with congruous and incongruous audiovisual stimuli consisting of the nonsense syllables /gig/, /gyg/, /geg/, and /gøg/, and asked to classify each stimulus as containing any one of 9 Swedish vowels. In nearly all cases, listeners perceived an unround auditory stimulus paired

with a rounded visual stimulus as a rounded vowel. For instance, auditory [eː] paired with visual [yː] was perceived as /øː/ in 127 of 128 cases, suggesting that visual cues have a strong effect on the perception of roundedness. On the other hand, when participants were presented with stimuli that were incongruous in terms of height, perception was dominated by the auditory channel. Traunmüller and Öhrström conclude that acoustic cues relating to rounding are relatively unreliable compared to those for height, causing listeners to attend to visual rounding cues even under ideal listening conditions. They note that these results support the "information reliability hypothesis," which states that, in multisensory perception, "perception is dominated by the modality that provides the more reliable information" (255).

In a second study, Traunmüller and Öhrström (2007b) investigated the perception of subphonemic contrasts in relation to audiovisual speech perception. Phonetically-trained listeners were presented with the same audiovisual stimuli used by Traunmüller and Öhrström (2007a) in video-only, audio-only, and audiovisual conditions. In one session, participants were asked to rate each stimulus they heard on the dimensions of lip rounding, lip spread, and position in a quadrilateral vowel space. In a second session, participants were asked to rate the vowels they *saw* along the same dimensions. The results of this experiment support the finding that listeners rely primarily on visual cues in judging vowel roundedness, while relying on auditory cues for judging height. Although participants in the video-only condition accurately rated stimuli in terms of rounding, ratings for height were less accurate, indicating the reliability of visual cues for perception of rounding but not for height. In addition, no significant difference was found between the roundedness ratings for the audio-only stimuli and for the audiovisual stimuli, suggesting that the acoustic cues for rounding offer little benefit for listeners. Finally, the roundedness ratings for the video-only and audiovisual stimuli showed a positive correlation, while the height ratings did not, suggesting that the addition of visual information had no effect on the perception of height.

More recently, a series of studies by Ménard et al. (2009), Ménard et al. (2013), Ménard et al. (2015), and Ménard et al. (2016) has demonstrated the importance of audiovisual speech perception to speech intelligibility through an investigation of differences in the use of visible articulation by sighted and blind speakers. Ménard et al. (2009) performed a study of the production and discrimination of vowel contrasts in sighted and congenitally blind adult speakers of Canadian French. In the perception experiment, participants completed an AXB task, in which they compared tokens along height, rounding, or backness continua between the vowel pairs /i/-/e/, /e/-/ɛ/, /ɛ/-/a/, /i/-/y/, and /y/-/u/. Overall, blind speakers showed greater auditory discrimination abilities than sighted speakers, as measured by peak discrimination score. In the production experiment, participants repeated each of the ten French oral vowels, and Euclidean distance was calculated for the same vowel pairs as used for the perception experiment. In addition, average vowel spacing (AVS; Lane et al. 2001) was calculated for each speaker, providing a measure of the overall vowel space. It was found that sighted speakers have a significantly larger vowel space than blind speakers, despite blind speakers' higher auditory discrimination. The authors interpret these results as an indication that the availability of visual speech cues influences speech production targets.

In subsequent research, Ménard et al. (2015) and Ménard et al. (2016) have investigated how blind and sighted speakers differ in the production of vowel contrasts in clear and conversational speech. They note that clear speech is characterized both by acoustic changes such as increased vowel duration and intensity, greater distance between vowel categories, and smaller distributions within vowel categories, as well as an increase in the magnitude of articulatory gestures. In order to examine whether the availability of visual speech cues influences speech production, they performed an articulatory and acoustic study of ten congenitally blind and ten sighted speakers of Canadian French. Crucially, they find that only sighted speakers produce larger lip movements in clear speech, while blind speakers rely on changes in tongue movement, a finding which suggests that the production of clear speech,

and articulation more generally, relies on achieving both articulatory and acoustic targets. In order to enhance intelligibility, sighted speakers seem to consider how their speech will be conveyed not only acoustically, but optically as well. On the other hand, for speakers who have never had access to vision, only clarity of the acoustic signal is relevant.

Despite the well-known influence of visual cues on speech perception, relatively few studies have considered the role of visual speech perception in sound change. One exception is a study by Johnson, DiCanio, and MacKenzie (2007), who investigate the role of visual cues in determining the place of articulation of excrescent nasals. They observe that in the variety of French spoken in Toulouse, words with a final nasal vowel like *savon* 'soap' are realized as [savɔŋ], rather than standard [savɔ̃]. They conducted a set of experiments to test the hypothesis that the tendency for nasalized vowels to alternate with velar nasals, rather than labial or coronal nasals, is due to both the visual and auditory (Ohala 1975) similarity between velar nasals and nasalized vowels. Participants were exposed to audio-only and audiovisual stimuli containing CVN sequences with a final [m], [n], [ŋ], or a 'placeless' nasal [x̃]. The placeless nasals were created from CVm tokens by removing the formant transitions leading into the nasal consonant, thereby removing the cues to the nasal's place of articulation. In one experiment, participants were asked to identify the place of articulation of the stimuli by labeling them with "m," "n," or "ng." In the audio-only conditions, participants performed roughly at chance (33%) in identifying [m], but correctly identified [n] and [ŋ] at higher than chance rates. Placeless nasals were most often identified as "ng," confirming their acoustic similarity to velar nasals. In the audiovisual and video-only conditions, however, participants correctly identified [m] in over 90% of cases, while [n] and [ŋ] were correctly identified at rates of approximately 80%. The auditorily placeless tokens were paired with video of each of the three nasals ([m n ŋ]), and listener identification generally matched the visual input. These results show that audiovisual input significantly influences listener identification of nasal place of articulation. In another experiment, they

more directly tested the visual similarity of velar nasals to nasalized vowels, both of which have no visible place of articulation. In this experiment, a new set of stimuli were used in which the CVN sequences were created from words ending in [m], [n], and [ŋ], as well as nonce CV syllables containing nasalized vowels, produced by an L2 speaker of French. The final section of the audio track for each stimulus was replaced with noise, in order to completely mask each nasal's place of articulation. The results show that audio-only [m] and [n] stimuli were typically identified as "n," while audio-only [ŋ] and [x̃] stimuli were typically identified as "ng." When the stimuli were presented audiovisually, identification of [m], [n], and [ŋ] was substantially more accurate, and stimuli containing [x̃] were more likely to be identified as "ng" than in the audio-only condition. Johnson and colleagues therefore argue that nasalized vowels are liable to be misperceived as velar nasals in diachronic sound change, because velar nasals are both visually and auditorily similar to nasalized vowels.

Other researchers have attributed asymmetries in sound change typology to variability in both the auditory and visual domains. McGuire and Babel (2012) investigate the common patterns of *th*-fronting, where /θ/ becomes /f/, and *th*-stopping, where /θ/ becomes /t/. /θ/ > /f/ changes have frequently been attributed to the acoustic similarity of /θ/ and /f/, but the authors note that an account based on acoustic similarity alone fails to predict the cross-linguistic rarity of /f/ > /θ/ changes. They argue that this asymmetry arises instead from inter- and intra-speaker variability in the production of /θ/, which can be produced either with dental or interdental articulations. They conducted three experiments, in which participants were presented with audio-only, video-only, or audiovisual stimuli consisting of syllables containing /f/ or /θ/, as produced by 10 separate talkers. They find that discrimination of /f/ and /θ/ is greater in the audiovisual condition than in the audio-only or video-only conditions. In addition, in the video-only condition, a strong positive correlation is observed between listener sensitivity to the /θ/-/f/ contrast and the degree to which a speaker's production is interdental. Listener sensitivity to the /θ/-/f/ contrast is greater for

speakers who produce interdental /θ/ than for speakers who produce /θ/ with no visible tongue gesture. These results suggest that the visibility of interdental articulations may play a role in maintaining the contrast between /θ/ and /f/, and that this contrast can sometimes fail due to variability in articulation.

Finally, Johnson (2015) tests the hypothesis that stop debuccalization can occur as the result of temporal misalignment of auditory and visual speech cues. To investigate this question, he considers debuccalization of stops in homorganic and heterorganic clusters. He notes that stops in homorganic clusters are more likely to debuccalize than those in heterorganic clusters, as in Hayu (Sino-Tibetan), where voiceless stops become a glottal stop when a nasal-initial suffix is added. This alternation is observed in (4):

(4) /dip-me/ → [diʔme]
/put-no/ → [puʔno]
/pʊk-ŋo/ → [pʊʔŋo]

Johnson hypothesizes that this pattern is an effect of reduced visual distinctiveness in homorganic as opposed to heterorganic clusters, due to a decrease in the amount of articulatory movement. This may lead listeners to attribute the visible articulatory movements to only one segment of the cluster, such that they interpret the other segment as a glottal stop. Such a change would be an instance of hypercorrection under Ohala's model of innocent misapprehension; additional examples of sound changes involving hypercorrection of lip rounding cues will be discussed in Chapter 5.

To test this hypothesis, Johnson (2015) conducted a study in which phonetically-trained listeners were asked to transcribe a series of stop+sonorant clusters. Each participant was presented with audiovisual stimuli consisting of 32 stop+nasal or stop+l clusters, of which 8 were homorganic, with the audio offset with respect to the video by −200, −100, 0, 100, or 200 ms. Participants were told that the stimuli were words of CVCCV structure, and were

asked to identify the second C as /p/, /t/, /k/, or /ʔ/. Johnson finds that perception of consonants in word medial clusters depends heavily on the place of articulation of the neighboring consonant. Overall, the highest rate of glottal stop responses was found for [t] stimuli, which is unsurprising given that *t*-glottalization actively occurs in American English. He finds that labials were most susceptible to misperception in homorganic clusters, and that labials were most affected by video asynchrony. When the audio signal preceded the video signal, listeners reported hearing [p] as [k] or [ʔ], but only when an [m] followed the stop. In these stimuli, the visual stop closure occurred after the audio stop closure, so listeners associated the lip closure with [m], rather than the stop. Johnson argues that, although the effect is small, audiovisual integration may provide a push toward debuccalization of labial stops, even if other factors are ultimately needed to account for the propagation of this sort of change.

1.2.2 Other Modes of Perception

In addition to visual speech cues, listeners have been shown to be sensitive to perceptual information across a number of other sensory modes. Fowler and Dekle (1991), for instance, demonstrate that listeners integrate sensations experienced haptically, even among untrained listeners. They conducted a set of experiments in order to determine whether cross-modal perception arises as a result of experience and association, or due to the direct perception of the source of a stimulus. Participants were exposed to one of two conditions. In the first condition, participants were exposed to auditory stimuli paired with matched or mismatched orthographic representation. In the second condition, participants heard the same auditory stimuli while placing their hand on a model's face in accordance with the Tadoma method of lipreading, with the model producing matched or mismatched lip movements timed with the onset of the auditory stimulus. They find that participants in the second condition reliably integrated the articulatory information perceived haptically, such that stimuli ('ba' and

'ga') were significantly more likely to be perceived as 'ba' when paired with a haptically-perceived lip closure. However, mismatch of the orthographic and heard stimuli did not result in misperception. Fowler and Dekle interpret these findings from the perspective of direct realism, arguing that the integration of haptic speech cues is a result of the perception of a cue's source, in this case vocal tract gestures, rather than an association of haptic and auditory cues in memory. This interpretation is made on the basis that felt and heard information are associated because they necessarily derive from a common environmental source, while spelled and heard information are associated only by convention, not as a necessary consequence.

Gick and Derrick (2009) performed an experiment investigating whether listeners integrate other types of tactile information in speech perception, in this case aerotactile information. In this experiment, participants were exposed to auditory stimuli containing the syllables /pa/, /ba/, /ta/, and /da/, in one of three conditions. In the first condition, a synthetic puff of air was applied to the listener's hand for 50 milliseconds prior to the onset of the vowel, similar to the timing of aspiration in English consonants. This type of sensation is relatively natural, they argue, as speakers likely feel their own breath on their hands on occasion during speech. In the second condition, the same puff of air was applied to the center of the listener's neck, in a location where speakers are unlikely to feel their own breath during speech. A third group of listeners were exposed to the auditory stimuli with puffs of air directed away from their skin, to ensure that the airflow itself was inaudible. In all conditions, participants heard the auditory stimuli both with and without a simultaneous puff of air. The results show that stops are significantly more likely to be perceived as aspirated when paired with a puff of air. For instance, listeners correctly perceived the syllable /da/ in nearly 90% of cases when heard without a puff of air, but performance drops to 70% when /da/ is heard with a puff of air. The addition of an air puff enhanced the perception of aspirated stops, which were correctly perceived 70% of the time with a puff of air, but only

60% without. While the effect of the air puff was stronger when applied to the hand than to the neck, there was no effect of hearing the air puff when airflow was directed away from the listener. These results suggest that listeners integrate aerotactile cues in speech perception, even when presented in a location where listeners would have little previous experience perceiving such cues.

Finally, in a demonstration of how listeners integrate even quite indirect perceptual cues, Mayer et al. (2013) examine the visual perception of airstream cues. They conducted an experiment in which listeners were exposed to stimuli containing both a speaker, who produced tokens of the words *pom* and *bomb*, as well as a candle, which flickered as a result of aspiration of /p/. Stimuli were presented in both congruous and incongrous audiovisual conditions, such that the candle did not flicker with audio of /b/, and did flicker with audio of /b/, respectively. Flickering of the candle significantly increased the likelihood of a token being perceived as *pom*, suggesting that listeners possess some awareness of the aerodynamic effects of aspiration on objects in the environment, and integrate this knowledge during speech perception.

1.2.3 Gestural Models of Perception

The findings that listeners are sensitive to non-auditory perceptual cues has been a source of debate in theories of speech perception. For proponents of gestural models of speech perception, the multimodal nature of speech perception is taken as evidence that the objects of perception are not the sounds themselves, but the real or intended articulatory gestures used to produce those sounds. One early gestural approach to speech perception is the Motor Theory of Speech Perception, developed by Liberman and colleagues at Haskins Laboratories in the latter half of the twentieth century. Motor theorists argue not only that listeners perceive gestures, not sounds, but also that speech perception involves the motor system

of the brain and occurs in a specialized, language-specific module of the mind. The origins of Motor Theory lie in the development of speech synthesis tools and the search for invariants in the acoustic speech signal (Liberman, Delattre, and Cooper 1952; Liberman 1957). It was observed through these efforts that the mapping from acoustic cue to phonetic category was both one-to-many and many-to-one. Contextual variation, for instance, is a one-to-many mapping, where the same acoustic cue can be mapped to multiple phonetic units depending on the surrounding context. As shown in Figure 1.1, Cooper et al. (1952) found that synthesized stop bursts varied in their identification depending on the frequency of the second formant of the following vowel. While all high frequency bursts are identified as /t/, lower frequency bursts may be identified as /p/ when they appear before a high vowel, but as /k/ before a low vowel. On the other hand, phonetic categories are associated not with a single, invariant acoustic parameter like burst frequency, but with many acoustic cues that combine and trade to give rise to a constant percept. For example, although laryngeal contrasts like that between /p/ and /b/ are represented phonologically as a difference in the value of [$\pm$voice], the acoustic correlates of this contrast are numerous, and include differences in voicing duration during closure, VOT duration, fundamental frequency, duration of stop closure, and formant transitions into the following vowel, among many others (Lisker 1978, 1986). If one of these cues is diminished in strength or deleted altogether, it can be substituted with another, while the percept remains constant. The interpretation of these findings provided by the Motor Theory is that the perception of a phonetic category does not rely on an invariant acoustic cue; as noted by Liberman et al. (1967), "there is typically a lack of correspondence between acoustic cue and perceived phoneme, and in all these cases it appears that perception mirrors articulation more closely than sound" (453). Listeners incorporate a variety of acoustic cues to obtain information about the source of these sounds, which is a vocal tract gesture. It is not the acoustic cues themselves that are the object of perception, but rather the gestures used to produce them. Given the central

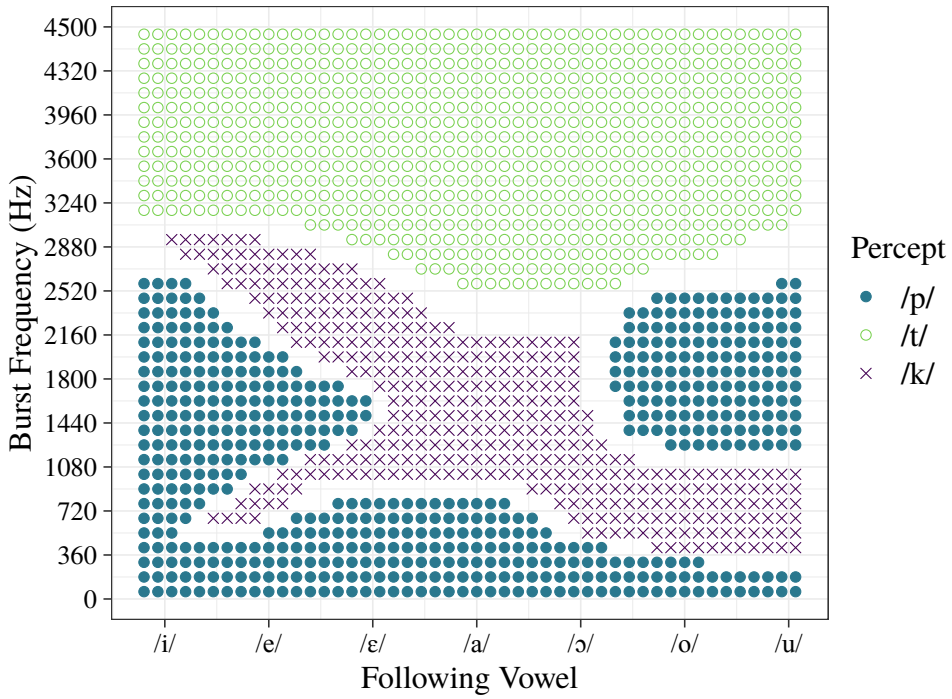


Figure 1.1: Identification of stop bursts of varying frequency in the context of seven English vowels. Adapted from Cooper et al. (1952).

role of articulation in gestural theories of speech perception, visual cues have served as an important source of evidence for the perception of gestures, as they provide listeners with direct evidence of the speaker's articulation. Fowler (1996), for instance, writes that "listeners perceive gestures, and some gestures are specified optically as well as acoustically" (1733).

1.2.4 Auditory Models of Perception

Galantucci, Fowler, and Turvey (2006) note that the support for gestural models of perception by linguists has been relatively limited, despite these models finding broader acceptance

in other areas of cognitive science and psychology. More widely accepted within linguistics are acoustic and auditory theories of speech, including Auditory Enhancement theory (Diehl and Kluender 1989) and Quantal theory (Stevens 1972, 1989). As discussed in Section 1.1, auditory perception also plays a crucial role in Lindblom's H&H theory, as well as Ohala-style models of sound change.

Diehl and Kluender (1989) argue against gestural models of speech perception, asserting instead that listeners are primarily attuned to the auditory rather than the articulatory properties of speech. They propose the Auditory Enhancement Hypothesis, which states that phonological contrasts are selected by speech communities on the basis of perceptual enhancement, such that acoustically more distinct segments and features are chosen. They provide several arguments in support of this hypothesis. First, they note that the vowel systems of the world's languages tend to be organized such that auditory distinctiveness is maximized (Liljencrants and Lindblom 1972; Lindblom 1986). This is evidenced by the fact the most common vowels are /i/, /u/, and /a/, which respectively exhibit maximal F2, minimal F2, and maximal F1. Although they note that these maxima could also be defined in articulatory terms, they suggest that the articulatory strategies used to produce these appear to be oriented toward maximizing auditory dispersion. Whereas the contrasts /i/-/u/ and /y/-/ɯ/ are equally dispersed in articulatory terms, only /i/-/u/ exhibits maximal acoustic dispersion. In addition, they argue that auditory enhancement is observed in the production of consonants and tone. For instance, they argue that the covariation of voicing correlates, such as closure duration, F0, and vowel length, is produced by speakers in order to maximize the strength of laryngeal contrasts.

Stevens (1972, 1989) considers distinctive features in terms of the relationship between articulation and acoustics, and proposes a quantal theory of speech production and perception. Stevens observes that when articulatory parameters, such as tongue backness, are manipulated, there is a nonlinear relationship between tongue position and the acoustic

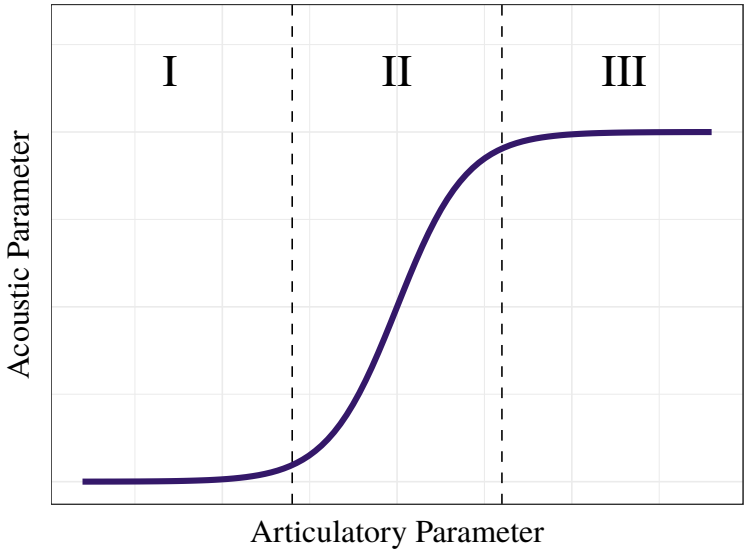


Figure 1.2: Nonlinear relationship between articulatory and acoustic parameters. Adapted from Stevens (1989).

signal. This relationship is schematized in Figure 1.2, where as the value for an articulatory parameter is manipulated, the value of an acoustic parameter changes in a sigmoidal fashion, with stable regions for high and low values of the articulatory parameter (regions I and III) and abrupt change at some critical point (region II). For instance, nasal segments are produced with a low velum, allowing air to pass through the nasal cavity. As the velum raises toward the pharyngeal wall, little acoustic change is observed as long as the velar port remains open (region I). When the velum makes contact with the pharyngeal wall (region II), nasal resonance abruptly ceases, although the velum may continue to raise, again with little effect on the acoustic signal (region III). Regions I and III are theorized to define opposing values for a given feature; in this example, region I defines [+nasal], while region III defines [−nasal], with region II serving as the boundary between these values.

1.2.5 Exemplar Models of Perception

Johnson (1997) proposes the use of exemplar models of perception as a means of explaining how listeners adapt to interspeaker variation, mapping clouds of individual productions to abstract phonetic categories. In exemplar models, phonetic categories are defined not as a singular abstraction, but rather as a complete set of previously-experienced tokens, or exemplars. Johnson suggests that perceived exemplars are stored with a variety of labels, including how and by whom an exemplar was produced. These labels include auditory properties, such as formant values, as well as speaker-specific information, including the speaker's name and gender. When new exemplars are perceived, they are compared to existing exemplars with respect to each of the stored labels. A category is activated when the percept is sufficiently similar to existing exemplars of that category. Exemplar models of speech perception are particularly relevant to the present investigation, as they are amenable to the incorporation not only of speaker-specific information, but also non-auditory perceptual cues and sociolinguistic information. In principle, exemplars may be defined not only in acoustic terms, but in articulatory terms as well. This would allow, for instance, speakers to select as production targets exemplars that are optimized for visual as well as auditory distinctiveness when speaking in a noisy environment.

1.3 Articulatory Variation

As noted above, the origins of the Motor Theory of Speech Perception lie in the "invariance problem," a search for some invariant property of speech, whether in acoustics or articulation. While the Motor Theorists argue that speech sounds more closely track their articulatory patterns than their acoustic cues, a wealth of evidence has demonstrated that speech sounds are highly variable in both their articulation and their acoustics. Perkell et al. (1993), for instance, examine the contributions of tongue-body raising and lip rounding

to the production of /u/. They hypothesize that these two gestures covary in order to reduce the amount of acoustic variation in the production of /u/, and used Electromagnetic articulometry (EMA) to measure the position of the tongue and lips during production of /u/. It was found that the parameters for lip-rounding and tongue raising are negatively correlated, indicating that as lip-rounding increases, tongue raising decreases, and vice-versa, although the effect in this experiment is somewhat weak. They find support for the hypothesis that gestural covariation serves to decrease acoustic variation, and suggest that the weak results may be specific to English /u/, which is relatively unconstrained in its acoustic realization compared to other vowels that inhabit more-crowded areas of the vowel space.

De Jong (1994) considers the relationship between rounding and backing in /u/ in terms of enhancement features. Enhancement features, proposed by Stevens, Keyser, and Kawasaki (1986), are part of a phonetically-based theory of underspecification, in which only certain distinctive features need be specified to create lexical contrast, while other, redundant features serve to enhance the acoustic signature of the distinctive features. Perhaps the most prominent example of a distinctive-redundant feature pair is that of [back] and [round] in English. While the English back vowels (with the exception of /ʌ/) may be specified for both [+back] and [+round], only one of these features is needed to form a contrast with the front, unround vowels. Because the [round] contrast between /ʌ/ and /ɔ/ is marginal in many dialects of American English, Stevens, Keyser, and Kawasaki argue that only [back] is needed to form contrasts, with [round] serving as a predictable enhancement feature. [round] may therefore be eliminated from the non-low back vowels' specifications, with rules filling in the [+round] feature by default. De Jong tests this hypothesis with an articulatory study of backing and rounding in American English. Because rounding is redundant, it is predicted that the articulation of rounding should be more variable than that of backing, given that [back] is necessary to form contrasts. He uses data from the X-ray microbeam database to analyze the productions of three speakers from the Upper

Midwest. Each speaker repeated a set of eight words, each containing /oʊ/, in three accentual conditions. De Jong finds that both backing and rounding are highly variable for these speakers, with backing and rounding existing in a compensatory relation. When the tongue is unable to achieve its backing target due to coarticulatory demands from neighboring alveolar segments, rounding allows speakers to achieve the same lowering of F2. However, speakers varied in the extent to which rounding compensated for backing. While two of the three speakers increased the degree of lip rounding in the face of reduced backing, the other speaker reduced lip rounding along with the reduction in backing.

Stone and Vatikiotis-Bateson (1995) examine compensatory articulatory maneuvers which occur as a result of coarticulation, rather than perturbations induced experimentally. They use a combination of ultrasound tongue imaging, electropalatography (EPG), and jaw height measurement to investigate how movement of the articulators for /i/ and /a/ differs when in the context of /l/ or /s/, which require quite different tongue shapes. They hypothesized that "lo-spec" articulators, those which do not produce the primary constriction for a sound, may move in unpredictable ways in order to assist "high-spec" articulators, which are responsible for producing the primary constriction, in producing the required acoustic output when contextual constraints are imposed by neighboring segments. Stone and Vatikiotis-Bateson find that when /i/ occurs in the context of a neighboring /l/, the height of the tongue (and thereby the airflow channel width) remain similar, despite the position of the jaw being lower. Ultrasound data reveal that this is achieved by an increase in the degree of ATR, which serves to increase the height of the tongue constriction in the face of a lower jaw movement. They argue that the tongue tip and root engage in a trading relation, which is required to maintain a consistent acoustic output.

Given that articulation exhibits contextual variation within the speech of individual speakers, it is unsurprising that articulation also varies across speakers. Articulatory research has demonstrated that in some cases, speakers have a choice with regard to

how a given sound is articulated, such as American English /ɹ/, which is known to exhibit a broad range of interspeaker variation in its articulation (Delattre and Freeman 1968). Such variation is possible, in part, because different articulatory gestures can have similar effects on the acoustic signal. For instance, both velar and labial constrictions have a lowering effect on F2 (Fant 1973), while nasalization can simulate a lowering of F1, which is typically associated with tongue-raising (Krakow et al. 1988). There also exist cases, however, where articulatory variation does not arise, even though it is hypothetically possible. One example is diachronic /u/-fronting in British English, where speakers have been found to achieve this change through tongue-fronting alone, even though an increase in F2 can also be achieved by an unrounding of the lips (Harrington, Kleber, and Reubold 2011).

The most widely-studied example of covert articulatory variation is perhaps that of English /ɹ/. In early work, there were thought to be two distinct variants of /ɹ/, bunched and retroflex. These two articulations were described by Uldall (1958), who performed acoustic and static palatographic analyses of these variants in her own speech. Uldall refers to bunched /ɹ/ as "molar," and describes the tongue shape as bunching toward the upper molars, with the tongue tip and blade pointing toward the bottom of the mouth. Retroflex /ɹ/, which she refers to as "tongue-tip," exhibits a raised tip, with "the front of the tongue held concave to the palate" (104). She describes these articulations as nearly indistinguishable acoustically, noting that both have similar formant structure below 2500 Hz, corresponding to F1-F3, but that bunched /ɹ/ has an additional formant at 4200 Hz, which retroflex /ɹ/ lacks. Finally, she describes the distribution of these variants in her own speech, where bunched /ɹ/ is used syllabically in /ɚ/ and after stressed vowels, while retroflex /ɹ/ occurs in utterance-initial position and following the alveolars /s ʃ t d/.

Delattre and Freeman (1968) provide a more extensive classification of tongue shapes for English /ɹ/. They use cineradiography (X-ray) to examine the articulation of /ɹ/ by 46 speakers, finding that 6 different tongue configurations are used in American English. These

six variants include a relatively open vowel-like articulation found in non-rhotic varieties, a bunched articulation with simultaneous constrictions between the dorsum and the palate and the root and the pharynx, a tip-up variant with a pharyngeal constriction and a tongue tip constriction at the post-alveolar region, and several intermediate shapes. For British English, they argue that there are two distinct variants of /ɹ/: one which is non-rhotic, and occurs in postvocalic position, and one with a retroflex alveolar constriction that occurs in prevocalic position. Despite the wide range of variation found in /ɹ/, however, they find no significant difference in the formant structure for any of the rhotic variants. More recent work by Espy-Wilson (2004) has found that this holds true for F1-F3, but that different articulatory configurations for /ɹ/ are associated with differences in the values of F4 and F5.

Twist et al. (2007) tested the perceptibility of variation in /ɹ/, noting that within speakers, multiple articulations for /ɹ/ might be used as allophones in predictable contexts. Participants included both native English speakers and native speakers of Mandarin. Mandarin speakers were included because Mandarin contains retroflex sounds, and it was predicted that these speakers might perform better in distinguishing retroflex variants of /ɹ/. Participants were presented with stimuli in which /ɹ/ varied both in word position (pre- or post-vocalic) and in articulation (bunched or retroflex). They find that both Mandarin and English speakers perform poorly in identification of these variants, and suggest that the acoustic differences between /ɹ/ variants may be sufficient for justifying allophony within an individual's own speech, but not sufficient for a pattern to arise within the community.

In contrast to /ɹ/ in American English, Lawson, Scobbie, and Stuart-Smith (2011) find that /ɹ/ variation in Scottish English is socially stratified. They collected ultrasound and video data on the realization of postvocalic /ɹ/, captured during both conversational and wordlist speech. Tongue gestures were classified as either tip up, in which the tongue surface is straight or concave; front up, in which the tongue surface is convex, but not bunched; front bunched, in which the tongue tip and blade are lower than the rest of the tongue; and

mid bunched, in which the tongue front is low, but the tongue body is raised toward the palate. /ɹ/ tokens were also auditorily classified by the strength of rhoticity, independent of the articulatory classification. It is found that working class speakers produce /ɹ/ primarily with tip-up and front-up tongue configurations, while middle class speakers almost entirely produce bunched variants. The articulatory findings were closely related to the acoustic findings, with the tip-up and front-up configurations of the working class speakers being classified as less auditorily rhotic than the bunched variants produced by the middle class speakers. While it seems that an acoustic distinction is necessary for stratification of articulatory gestures to arise, there remains much work to be done before the relationship between acoustic variation and articulatory variation can be fully understood.

In addition to /ɹ/, other alveolar segments exhibit widespread variation in English. Bladon and Nolan (1977) performed an X-ray study of alveolar articulations, which had been argued to have both apical and laminal constrictions, in a variety of coarticulatory environments. In apical, or tongue tip, articulations, the tongue tip is pointed toward the alveolar ridge, while in laminal, or tongue blade, articulations, the tongue tip is located behind the lower teeth, with constriction produced by the tongue blade. Bladon and Nolan find that a simple two-way classification of articulatory shapes is insufficient, and propose a system based on the positions of the tongue tip and blade. In their system, tongue shapes are classified using two numbers, with the first digit referring to the height of the tongue tip and the second referring to the height of the tongue blade. The scale for each digit ranges from 1 to 3, with larger numbers indicating a position higher in the mouth. Thus, tongue types 3-1, 3-2, and 3-3 correspond to what have been referred to as apical articulations, since the tongue tip for all three types has a high position close to the alveolar ridge. The tongue blade can be positioned below the tongue tip (Type 3-1), such that the tongue surface is concave, it can be raised to the palate (Type 3-3), such that the tongue surface is convex, or it can be in-between (Type 3-2), such that the tongue surface is flat. Types 2-3 and 1-3 correspond

to laminal articulations, with the only constriction being produced by the tongue blade. For these types, the tongue tip can be positioned above (Type 2-3) or below (Type 1-3) the lower teeth. In their study, seven of the eight speakers produced laminal constrictions (Types 2-3 and 3-3) for /s z/, which they note contradicts many previous descriptions of the English sibilants as apical. Variation was greater in their sample for the tongue positions of /n l t d/, suggesting that /s z/ are somewhat resistant to coarticulation.

Finally, an ultrasound study of articulatory variation of particular relevance to the present investigation is reported by De Decker and Nycz (2012), who examine the tense [æ] system characteristic of Mid-Atlantic dialects. Their study is motivated by competing articulatory sources for phonetic distinctions, in this case the effects of nasality and tongue-position on the realization of tense [æ]. As noted above, nasalization has a lowering effect on F1, which can simulate tongue raising (Krakow et al. 1988). De Decker and Nycz therefore predict that speakers have the option of using nasalization to contrast between tense and lax [æ]. They find that at least three variations in articulatory strategy exist; one in which a three-way distinction in tongue position is associated with a three-way distinction in the acoustic signal, one in which a two-way distinction in tongue position is associated with a two-way distinction in the acoustic signal, and one in which a two-way distinction in the acoustic signal is associated with only a single lingual gesture. In the latter case, speakers seem to rely on nasalization to achieve the acoustic contrast. As discussed in section 1.2.1, visual speech perception cues are known to influence speech perception. If visual cues do play a strong role in governing articulatory variation, this would suggest that a greater amount of variation should be possible when nasalization is involved, due to the fact that movement of the velum is not visible to the listener.

1.4 Dissertation Overview

Research on sound change and variation has primarily focused on the acoustic and auditory properties of speech. For instance, misperception (as well as mis-/re-interpretation) of the acoustic signal plays a fundamental role in theories of sound change that follow in the tradition of Ohala (1981, 1983, 1989, 1993). In teleological theories of phonology (Lindblom 1990; Hayes, Kirchner, and Steriade 2004), speakers optimize articulatory effort and perceptibility of the *acoustic* signal. Stevens (1989) argues that there exist ranges where changes in articulation have little effect on the acoustic output; languages exploit these regions to allow for consistent acoustic output with less articulatory precision. Finally, most work in sociophonetics (until recently) has described sound changes in terms of their acoustic characteristics, not the articulatory mechanisms used to produce the change. Lawson et al. (2010) write that articulatory study in sociolinguistics is "a closed book."

However, it is well known that speech perception is influenced by a number of non-auditory sensory domains, including visual (Sumby and Pollack 1954; McGurk and MacDonald 1976; Mayer et al. 2013; Ménard et al. 2016), somatosensory (Houde and Jordan 1998; Nasir and Ostry 2006; Jones and Munhall 2005), and haptic perception (Fowler and Dekle 1991; Gick and Derrick 2009). In this dissertation, it is argued that considering the role of visual cues in speech perception can improve our understanding of two areas of phonological variation and change. First, in articulation, the range of possible variation may be restricted by the availability of visual speech cues to language learners. Second, in sound change, visual speech cues can inhibit misperception of the speech signal in cases where two sounds are acoustically similar.

This dissertation is organized as follows. Chapter 2 presents an experiment examining how the fronting of the back vowels /u o ʊ/ is achieved in two dialects of American English. It is shown that the observed increase in F2 in these dialects is primarily the result of

fronting of the tongue, rather than of an unrounding of the lips. Chapter 3 investigates another notable case of back vowel fronting, the fronting of /ɑ/ and /ɔ/ associated with the Northern Cities Shift. This chapter focuses specifically on how contrast is maintained between the two vowels in terms of articulation; while some speakers produce a relatively weak acoustic contrast between /ɑ/ and /ɔ/, the majority of speakers distinguish between the two vowels through a difference in lip rounding, with the lingual distinction sometimes being lost. Moreover, the majority of speakers enhance the lip rounding distinction between /ɑ/ and /ɔ/ in careful speech, optimizing their articulatory patterns for both auditory and visual perceptibility. Chapter 4 describes the results of an audiovisual perception experiment investigating the role of visual lip rounding cues in enforcing the use of lip rounding to contrast /ɔ/ from /ɑ/. It is shown that unround variants of /ɔ/ are more likely to be misperceived as /ɑ/ than round variants, making articulatory strategies in which /ɔ/ retains its rounding perceptually stronger than those where rounding is lost. Chapter 5 presents a review of historical sound changes involving labial segments, providing support for the hypothesis that visual speech perception cues can inhibit misperception-based sound change. Chapter 6 considers the implications of audiovisual speech perception for theories of phonology and sound change, and provides concluding remarks.

Chapter 2

Articulation of Fronted Back Vowels in American English

Work in variationist sociolinguistics has provided a wealth of data on the nature of linguistic variation and change, enhancing our understanding of how sound patterns vary between speakers, how sounds change over time, and how speakers use linguistic variation to construct social meaning. Despite the progress made in this field over the past several decades, however, little is known about the articulatory patterns that underlie sound changes, and whether these patterns vary between speakers. As discussed in Section 1.3, the existence of articulatory trading and variation suggests the possibility that learners acquiring a variable and changing linguistic form will adopt an articulatory configuration that differs from that of the previous generation, or from that of other members of the speaker's community. One sound change in particular, back vowel fronting, provides an interesting case for testing the factors that constrain articulatory variation. On a purely acoustic/auditory basis, it is predicted that any articulation that achieves the desired acoustic output should be as good as any other. This may be the case for sounds like tense [æ], which can be produced either through tongue raising or nasalization (De Decker and Nycz 2012), neither of which is visible to the listener. For labial segments, however, visual considerations may also play a role. In the case of back vowel fronting, an increase in F2 can be achieved either through tongue fronting or lip unrounding. While these strategies are equally good in terms of acoustics, they differ in their visibility. A speaker aiming to optimize the perceptibility of the /i/-/u/ contrast might therefore prefer articulatory configurations which maintain visible lip rounding.

This chapter presents the results of a study investigating the articulation of fronted back vowels in two dialects of American English. It is found that speakers from both Southern California and South Carolina achieve the acoustic fronting of the vowels /u/ and /o/ through fronting of the tongue rather than unrounding of the lips. While both strategies are predicted to be possible on purely acoustic grounds, it is argued that the strategy of tongue fronting not only offers greater articulatory ease, but also enhanced visual contrast with the front unround vowels.

2.1 Back Vowel Fronting in English

Labov (1994) proposes three general principles governing vowel shifts, based on evidence from a number of chain shifts observed among the world's languages. These include Principle I, the generalization that tense vowels tend to rise, and Principle II, that lax vowels tend to fall. Most relevant to the present discussion, Principle III states that, "in chain shifts, back vowels move to the front" (116). This tendency has been observed in numerous dialects of English, including British (Harrington, Kleber, and Reubold 2008), Australian (Cox 1999; Cox and Palethorpe 2001), New Zealand (Gordon et al. 2004), South African (Mesthrie 2010), and North American varieties (Labov, Ash, and Boberg 2006). Outside of English, back vowel fronting has been observed in French, where the vowel /y/ arose historically from the fronting of /u/ (Calabrese 2000), as well as in Swedish, Albanian, and Akha (Sino-Tibetan), among many other languages (Labov 1994).

In American and Canadian English, back vowel fronting is widespread, and can be found in nearly all parts of North America. The *Atlas of North American English* (Labov, Ash, and Boberg 2006) provides an overview of North American back vowel fronting, including the

fronting of /u/, /o/, and the nucleus of /aʊ/.[1] For /u/, Labov, Ash, and Boberg find a three-way split in terms of F2. In pre-lateral environments, /u/ remains in the high back region of the vowel space, with a low F2. Following a coronal onset, /u/ is strongly fronted, exhibiting a mean F2 of 1811 Hz. Finally, following non-coronal onsets, /u/ is fronted marginally, with a mean F2 of 1433 Hz. The strongest /u/-fronting is found in the Midland (in Kansas City, St. Louis, and Indianapolis), as well as in Toronto. In each of these particular cities, the mean F2 for /u/ exceeds 2000 Hz.

In contrast to the fronting of /u/, the fronting of /o/ is less strongly conditioned by the place of the onset consonant, such that /o/ is fronted only marginally following coronals. More generally, Labov, Ash, and Boberg (2006) find that /o/-fronting lags behind /u/-fronting throughout North American dialects. While the most fronted tokens of /u/ approach [y], the most fronted tokens of /o/ are typically not fronted beyond the center of the vowel space. Moreover, the fronting of /u/ and /o/ do not necessarily occur in tandem; some of the regions with the strongest degree of /u/-fronting do not exhibit extreme fronting of /o/. The Midland cities and Toronto, for instance, show only slight fronting of /o/.

Figure 2.1 presents an overview of back vowel fronting in North America, with the F2 for /o/ plotted against that of post-coronal /u/. As Labov, Ash, and Boberg (2006) explain, this figure reveals three patterns of back vowel fronting in North American English. While some areas, such as the Southeast, exhibit strong fronting of both /u/ and /o/, other regions, such as Canada, front only /u/. In Eastern New England, neither /u/ nor /o/ are fronted. Two regions where back vowel fronting is of particular interest are California,[2]where back vowel fronting is a stereotypical component of the California Vowel Shift, and the Southeast, where

1. Throughout this dissertation, vowels are referred to by their canonical General American transcription in IPA, rather than the word classes used by Labov (1994) or the lexical sets proposed by Wells (1982). Here, /u/ refers to (uw)/GOOSE, /o/ refers to (ow)/GOAT, and /aʊ/ refers to (aw)/MOUTH.

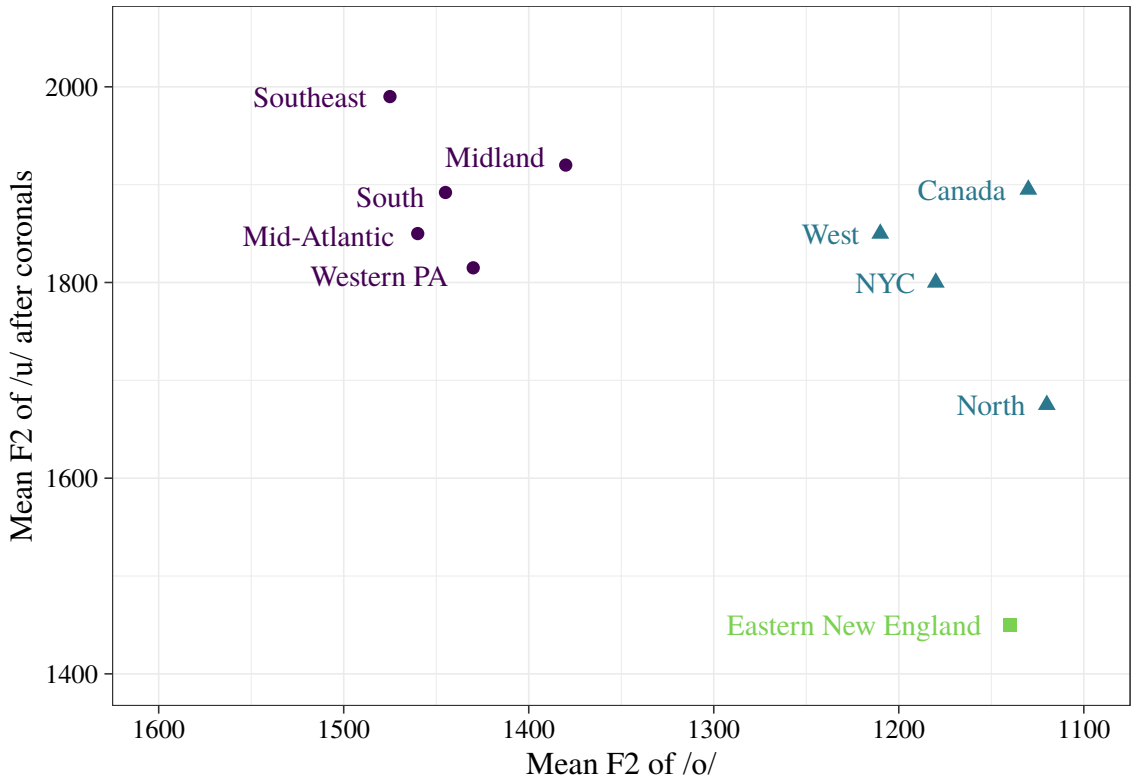


Figure 2.1: Mean F2 for post-coronal /u/ vs. mean F2 for /o/ in North American English. Adapted from Labov, Ash, and Boberg (2006, 157).

the fronting of both /u/ and /o/ is especially advanced. This chapter focuses on back vowel fronting in these two regions, the vowel systems of which are described in turn below.

2.1.1 The Southern Shift

Labov, Ash, and Boberg (2006) describe the speech of the American South as one of the most widely recognized dialects of North American English, noting that "Southern" is nearly always identified as a distinct dialect region in studies of dialect geography (e.g., Preston 1986, 1988). Southern American English exhibits a number of unique characteristics, including not only several vowel shifts, but also distinctive mergers and retention of phonemic contrasts not found in the rest of North America. For instance, the South is conservative in its maintenance of the /ʍ/-/w/ contrast, as in *which* vs. *witch*, and of the /iw/-/uw/ contrast, as in *dew* vs. *do*. Both contrasts have otherwise been lost in nearly all of North America. The South is also unique in exhibiting the *pin-pen* merger, such that the contrast between /ɪ/ and /ɛ/ is lost in prenasal contexts. Labov, Ash, and Boberg (2006) delineate the Southern dialect region, however, by the Southern Shift, one of two major ongoing chain shifts (along with the Northern Cities Shift) in North American English.

The Southern Shift, presented in Figure 2.2, is a combination of three sets of vowel changes, of which only the front chain shift is unique to the South but which in combination form a distinctive vowel system. The first stage of the Southern Shift is the deletion of the glide in the diphthong /aɪ/, in words such as *my* and *time*, which are pronounced as [maː] and [taːm]. Following the monophthongization of /aɪ/, the vowel /eɪ/, as in *made*, lowers to the previous position of the /aɪ/ nucleus, while /ɛ/, as in *bed*, raises to the position of /eɪ/, such that these vowels are reversed. In the third stage of the shift, /i/ and /ɪ/ undergo a similar reversal. The most widespread of these changes is the monophthongization of /aɪ/, while the reversal of /i/ and /ɪ/ is found only in a concentrated area, primarily in Alabama and eastern

2. In Figure 2.1, California is subsumed under the label "West."

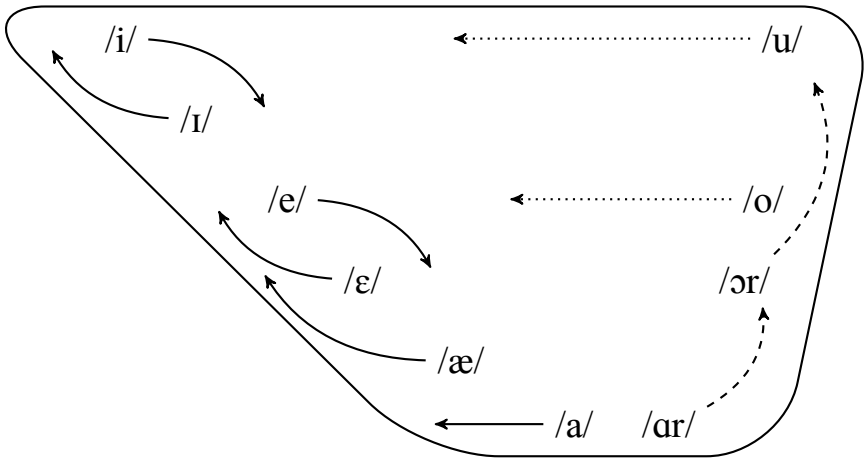


Figure 2.2: Schematic diagram of the Southern Vowel Shift. Solid lines indicate the Southern Shift, dashed lines indicate the back chain shift, and dotted lines indicate back vowel fronting. Adapted from Labov, Ash, and Boberg (2006, 242-44).

Tennessee. In addition to the shifting of the front vowels, the Southern Shift involves an upward chain shift among the back vowels, which Labov, Ash, and Boberg (2006) identify as the most common chain shift found in the world's languages. In the South, this chain shift occurs only before /ɹ/ and results in /ɔɹ/, as in *north*, being raised to the position of /oʊɹ/, as in *force*. In addition, /aɹ/, as in *start*, is raised to the former position of /ɔɹ/. Most relevant for the present study is the fronting of the back vowels /u/ and /o/. As noted above, this vowel shift is widely observed throughout North America, and not unique to Southern American English. However, the South is particularly advanced in its fronting of /o/ and is also among the most advanced regions of the United States in its fronting of /u/. In addition, the South exhibits some degree of fronting of /o/ and /u/ before /l/, where these vowels typically do not undergo fronting due to the velarization of /l/ in coda position.

Extreme fronting of /o/ in the South was observed by Thomas (1989), who describes a study of /o/-fronting in Wilmington, North Carolina. Thomas analyzes recordings created

in 1973 and 1974 of 19 older (55+) black and white women and 28 younger (<30) white and black men, scoring tokens of /o/ in terms of their degree of frontedness. Scores are assigned on a scale from 0–5, with 0 representing unfronted tokens realized as [oː] and 5 representing fronted tokens realized as [ɜʏ]. He finds that older white speakers exhibit a mean frontedness rating of 2.61, while younger white speakers exhibit a mean rating of 4.11. However, there is no significance in the degree of frontedness by older and younger black speakers, who exhibit mean frontedness ratings of 1.25 and 1.52 respectively, a difference he attributes to the then-recent desegregation in North Carolina. Thomas notes three regions from which the fronting of /o/ appeared to be spreading: North Carolina, the Delmarva peninsula, and the region along the Georgia-Alabama border. However, the pattern of /o/-fronting in the South has since become widespread, with the South exhibiting the highest average degree of /o/-fronting among the regions studied in the *ANAE* (Labov, Ash, and Boberg 2006).

One Southern city with a notable pattern of back vowel fronting is Charleston, South Carolina, where Baranowski (2008) performed a sociophonetic analysis of 43 native Charlestonians. He finds that for middle and upper class speakers, /u/ is especially advanced after coronals, with a mean F2 upwards of 2080 Hz. In contrast to most other dialects of American English, /u/ is highly fronted even after non-coronal consonants, with a mean F2 of at least 2000 Hz for middle and upper class speakers, and as high as 2200 Hz for upper middle class speakers. With regard to the fronting of /o/, the Charleston dialect is also quite advanced, reaching an F2 of over 1800 Hz for the youngest speakers. Most areas outside the South and Midland exhibit a mean F2 for /o/ of 1200 Hz or less (Labov, Ash, and Boberg 2006). The extreme parallel fronting of /o/ and /u/ found in Charleston is notable, given that the Midland cities with the highest rate of /u/-fronting (Kansas City and Indianapolis) do not strongly front /o/ (Labov, Ash, and Boberg 2006). Moreover, while most speakers exhibit a greater degree of /u/-fronting after coronal consonants than after non-coronal consonants, the speakers studied in Baranowski's study exhibit nearly the same degree of fronting in

both environments. This finding is corroborated by the Southern speakers analyzed in the *ANAE*, who exhibit a mean F2 for /u/ nearly as high as that of Midland speakers, but who show a much lower effect of coronal onset.

This pattern of back vowel fronting raises several interesting questions with respect to articulation. Because these speakers front /u/ and /o/ in both coronal and non-coronal contexts, as well as before /l/, the association of back vowel fronting with coronal coarticulation may be lost. In regions where /u/ is fronted only in contexts with a coronal onset and non-lateral coda, fronted /u/ can be analyzed as a predictable allophone of backed /u/ that appears due to the anterior tongue position of the preceding consonant. For speakers in those regions, fronting /u/ by unrounding the lips is not necessarily predicted, as this strategy would result in an increase in F2 even in contexts where the tongue remains backed.[3] It would be possible to achieve this allophonic patterning by unrounding the lips for /u/ only when it follows a coronal onset, but such a pattern of articulation would be unmotivated by coarticulatory pressure. However, for speakers for whom /u/ and /o/ are fronted in all phonological environments, either tongue fronting or lip unrounding might provide sufficient increases in F2. As such, speakers may have more freedom to produce /u/ and /o/ with backed tongues and unround lips than speakers for whom back vowel fronting is the result of coarticulation from coronal onsets.

2.1.2 The California Vowel Shift

California English is characterized by the California Vowel Shift (CVS; Eckert 2008; Hall-Lew 2009; Podesva 2011), a counterclockwise rotation of the front and low vowels accompanied by the fronting of /u/, /ʊ/, and /o/. This pattern is presented in Figure 2.3. Unlike the Northern Cities Shift and Southern Shift, the CVS is not typically considered to constitute a chain shift; the term instead describes a set of characteristic changes that, in combination,

3. For example, in post-velar contexts, where the F2 for /u/ typically remains relatively low.

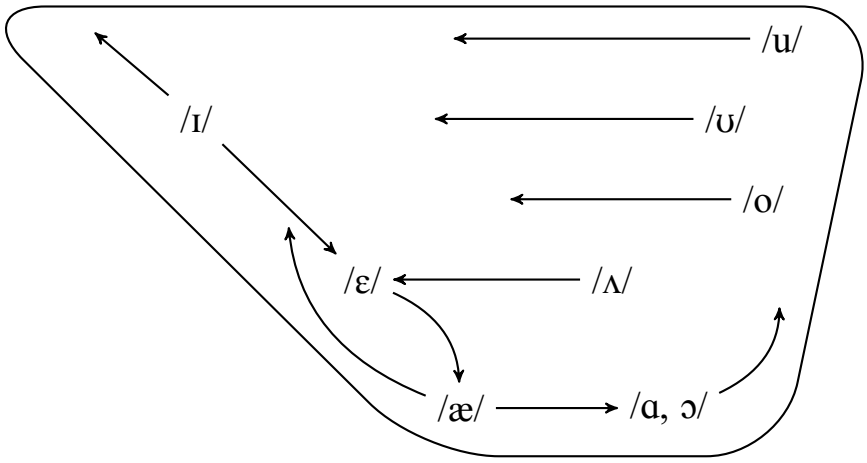


Figure 2.3: Schematic diagram of the California Vowel Shift. Adapted from Eckert (2008).

define the vowel system of California English. While the CVS has begun to receive attention as a discrete phenomenon only recently, individual components of the shift have been observed for several decades.

One of the earliest descriptions of the CVS comes from a series of studies conducted by Hinton et al. (1987), who analyzed ongoing vowel changes in speakers from Northern California, particularly the San Francisco Bay Area. The focus of their study was to compare the California English of the late 1980s to descriptions of California English from the *Linguistic Atlas of the Pacific Coast* (Reed and Metcalf 1952). They note that California was not widely recognized as a distinct dialect region in the 1950s or 1960s, and that linguistic studies from the time focused on the East Coast origins of Californian speech patterns. However, recognition of California as a distinct dialect region had begun by the 1980s (Preston 1986), by which time Californian speech had become the subject of parody in popular media.[4] Among

4. Hinton et al. (1987) mention three popular 1980s parodies, including the Frank Zappa song "Valley Girl," Whoopi Goldberg's "Surfer Chick" character, and the novel *The Serial*. More recently, Californian speech has been parodied in the *Saturday Night Live* skit "The Californians," which was

the changes analyzed by Hinton et al. (1987) are the backing of /æ/, the lowering of /ɪ/ and /ɛ/, and the fronting of /o/, /ʊ/, and /u/. Hinton et al. suggest that the most noticeable of these changes is the fronting of the mid and high back vowels, and find a high degree of frontedness for /o/, particularly by speakers between the ages of 16-22. Older speakers, however, showed little /o/-fronting, suggesting that the fronting of /o/ is a recent change. This is supported the results of Reed and Metcalf (1952), who found that /u/ was fronted minimally only by a small subset of speakers, and /o/ was not fronted at all.

Another early study of vowel fronting in California English comes from Hagiwara (1995, 1997). Hagiwara performed an acoustic analysis of 11 monophthongal vowels as produced by Southern California speakers, comparing the findings to those of Peterson and Barney (1952), who analyzed Mid-Atlantic speakers, and Hillenbrand et al. (1994), who analyzed speakers from the Upper Midwest. Participants produced the vowels /i ɪ e ɛ æ u ʊ o a ʌ ɝ/ in hVd, tVk, and bVt syllables. Hagiwara finds that Southern California speakers exhibit a mean F2 for /u/ of 1500 Hz for women and 1300 Hz for men, in contrast to Hillenbrand et al. (1994), who found a mean F2 for /u/ of close to 1000 Hz and Peterson and Barney (1952), who found a mean F2 for /u/ of less than 900 Hz for men and less than 1000 Hz for women.

2.1.3 Articulation of Fronted Back Vowels

Despite the attention received by back vowel fronting in the literature, little is known about the articulatory realization of fronted back vowels, given that work on variation in back vowel fronting has focused almost exclusively on acoustic data. However, because an increase in F2 can be the result of any gesture which shortens the front cavity of the vocal tract, including tongue fronting or lip unrounding, it cannot be known a priori which of

the subject of investigation by Pratt and D'Onofrio (2017). They analyze the role of jaw setting in the imitation of Californian speech and discuss more generally the portrayal of California speech styles in popular media.

these strategies is actually chosen by speakers in producing fronted /u/ and /o/. As noted in Section 1.3, studies by de Jong (1994), Perkell et al. (1993), and others have observed compensatory relationships between backing and rounding, and individual variation in the extent to which speakers rely on these articulatory gestures. However, these studies are limited to small sample sizes and do not specifically analyze the articulation of back vowels in speakers of differing dialects. While some sociolinguistic studies do describe the articulation of back vowel fronting, instrumental articulatory data on vowel shifts is generally lacking, and articulatory descriptions are instead based on inferences from acoustic data or on impressionistic observations. Moreover, descriptions and transcriptions of back vowel fronting often conflict with one another. For instance, Hinton et al. (1987) write that California back vowels "are clearly more front and less rounded than their 1950s counterparts" (119) and Hagiwara (1997) suggests they are "typically unrounded" (657). Thomas (2001), on the other hand, writes that "some authors have asserted that /u/ is undergoing unrounding as it is fronted, but I am skeptical about that" (34). Eckert (2008) suggests there may be variation in both the articulation and acoustic realizations of fronted back vowels, describing two types of fronting in California: "Surfer" and "Valley Girl." She transcribes the Surfer variant as [y], as in [dyd] *dude*, suggesting that this variant is fronted, but not unrounded. On the other hand, the Valley Girl variant exhibits both fronting and unrounding of the nucleus, such that *food* is realized as [fɪwd] and *goes* is realized as [gɐwz]. De Jong (1994) similarly argues that variation in rounding exists among back vowels in American English, such that speakers from Southern California exhibit "little or no rounding" on non-low back vowels, while southern Midwest (or Midland) speakers "seem to be losing the backing contrast" (70). Although de Jong's study does investigate the presence or absence of tongue and lip gestures in the articulation of /u/, he analyzes only speakers from the Upper Midwest, where /u/ and /o/ typically do not undergo advanced fronting.

While the articulatory patterns underlying back vowel fronting in dialects of American English have gone largely unstudied, a small number of studies demonstrate that fronted back vowels in dialects of British English are produced with a fronted tongue and rounded lips. In other words, the fronted back vowels are indeed fronted, not unrounded. Harrington, Kleber, and Reubold (2011) performed a set of articulatory and perceptual experiments investigating the role of the tongue and lips in the fronting of /u/ in Standard Southern British English (SSBE). They find that speakers of SSBE uniformly produce fronted /u/ with rounded lips and a fronted tongue. In the first experiment, the degree of anticipatory lip-rounding on /s/ was measured acoustically, with the hypothesis that if fronted /u/ in SSBE remains round, a greater degree of coarticulatory rounding will be observed in words like *soup* when compared to those like *seep*. The experimental results confirm this hypothesis, as demonstrated by a significantly lower spectral peak for the /s/ in *soup* than for that in *seep*. Moreover, the spectral differences between *soup* and *seep* are similar for both younger and older speakers, suggesting that the degree of lip-rounding for /u/ has not decreased over time. In the second experiment, native German speakers were presented with video (sans audio) of English /i/, /u/, and /ɔ/, and asked to classify these tokens as /i/, /y/, /u/, or /o/. Participants were also asked to classify the same visual stimuli cross-dubbed with audio of English /i/. For the video-only stimuli, at least 97.5% of /u/ and /ɔ/ tokens were classified as a round vowel (i.e, /u/, /y/, or /o/), suggesting that lip rounding in fronted /u/ is detectable, and not confusable with /i/. For the audiovisual stimuli, congruous /i/ stimuli were classified as /i/ in nearly 100% of cases. On the other hand, incongruous stimuli containing auditory /i/ paired with visual /u/ or /ɔ/ were classified as /u/ or /ɔ/ more often than they were classified as /i/. Finally, Harrington and colleagues performed an EMA analysis of five SSBE speakers, in which the tongue position and degree of lip protrusion for /u/ and /ʊ/ were compared to that of /i/, /ɪ/, /ɔ/, and /ɒ/. It was found that /u/ was closer to /ɔ/ than to /i/ in terms of lip-rounding for all speakers. In addition, the tongue position for /u/ was closer to that for /i/ and /ɪ/ than

to any other vowel, while /ʊ/ exhibited a tongue position between those for the high front vowels and those for /ɔ/ and /ɒ/.

These findings suggest that the diachronic fronting of /u/ in SSBE has not involved an unrounding of the lips, but rather has consisted primarily of a repositioning of the tongue. Harrington, Kleber, and Reubold argue that this fronting of the tongue position for /u/ is the result of lingual coarticulation following coronal consonants, and that this sort of coarticulatory effect may be the source of Labov's (1994) Principle III of chain-shifting. They note that as a result of these changes, lip-rounding has become the primary means through which /u/ is contrasted from /i/.

In a study of back vowel fronting in Scottish English, Scobbie, Lawson, and Stuart-Smith (2012) show that fronted /u/[5] is produced with a tongue position that is fronted, but also lowered. This additional finding of tongue lowering, such that the height of the tongue for /u/ is lower than that for /e/ and /ɛ/, sets the Scottish fronted /u/ apart from fronted /u/ in other varieties of British English, and results in an acoustic realization close to [ŏ] or [ʉ̟]. A cross-dialectal study of /u/ fronting by Lawson, Stuart-Smith, and Mills (2017) confirms the articulatory configuration for fronted /u/ found by Scobbie, Lawson, and Stuart-Smith (2012), and shows that the maximum tongue height for /u/ in Scottish English is indeed lower than the maximum tongue height for Anglo and Irish varieties of English. This finding demonstrates that the processes of /u/ fronting throughout English are not monolithic.

2.2 This Experiment

Although /u/-fronting is found in many varieties of English, it is unknown to what extent the findings of Harrington, Kleber, and Reubold (2011), Scobbie, Lawson, and Stuart-Smith (2012), and Lawson, Stuart-Smith, and Mills (2017) extend to other varieties, particularly

5. Scottish English /u/ encompasses the lexical sets goose and foot (Wells 1982), the latter of which is produced in American English with a lax vowel, [ʊ].

American English. While descriptive (Thomas 2001) and articulatory data (de Jong 1994) for American English /u/ and /o/ do exist, those data are limited in scope. This chapter presents a study of /u/-fronting and /o/-fronting in two dialects of American English. The goal of this experiment is to determine whether back vowel fronting in American English is achieved by tongue fronting, lip unrounding, or some combination of the two. While it is hypothetically possible to achieve /u/-fronting via unrounding of the lips, this strategy may be dispreferred if it results in a less-perceptible contrast between /i/ and /u/ or /ɪ/ and /u/, a possibility that is suggested by the finding that visual cues contribute to the accurate perception of vowel rounding (Traunmüller and Öhrström 2007a, 2007b; Valkenier et al. 2012).

A second goal of this study is to investigate the extent to which the processes of back vowel fronting in California and South Carolina differ from one another in their underlying articulatory strategy. With respect to the parallel fronting of both /u/ and /o/, the variety of English spoken in the South is among the most advanced in North America. In contrast, the California variety of English exhibits strong fronting of /u/, but does not necessarily show advanced fronting of /o/. Acoustic studies have also shown that these varieties differ in the phonological conditioning of vowel fronting. While vowel fronting in California English is strongly predicted by the presence of a coronal onset, the effect of onset place in the South is much weaker, with coronal onsets favoring fronting only marginally. In addition, some speakers from the South produce fronted back vowels even before laterals, which is not observed in California (or most other parts of North America). As noted above, these differences in acoustic patterning might also be associated with differences in articulatory strategy. For speakers from California, the strategy of vowel unrounding is not predicted on coarticulatory grounds, because if /u/ were produced with unround lips, it should also show a raised F2 following velar or glottal onsets, which is not the case. On the other hand, because vowel fronting in South Carolina is not necessarily tied to its coarticulatory source,

the strategy of lip unrounding is predicted to be more likely (although it is not a necessary consequence).

2.3 Methods

2.3.1 Language Varieties

This chapter considers back vowel fronting in two varieties of American English. The first is that of Southern California, where back vowel fronting is a well known and stereotypical component of the California Vowel Shift. The second is that of South Carolina. Back vowel fronting in South Carolina, and in the Southeast more generally, has received attention in the literature for being especially advanced, particularly in Charleston, as reported by Baranowski (2006, 2008).

2.3.2 Participants

Twenty-five participants (9 men, 16 women) took part in the study, which was conducted at the University of South Carolina in Columbia, South Carolina and at the University of California, San Diego in La Jolla, California. Participants were required to be natives of South Carolina or of coastal Southern California, having been born and raised in their respective region at least through the age of 18. None of the participants from Southern California have lived outside the region, while five of the participants from South Carolina have lived outside South Carolina. Demographic information is presented in Table 2.1 for participants from Southern California and in Table 2.2 for participants from South Carolina. Fifteen speakers (7 men, 8 women) from Southern California took part in the study, with an age range of 18 to 34 years ($M = 21.5$, $SD = 4.7$). Ten speakers (2 men, 8 women) were from South Carolina, and ranged in age from 18 to 50 years ($M = 27$, $SD = 8.9$). All participants reported normal hearing and speech.

Table 2.1: Demographic information for Southern California participants. *SoCal Origin* indicates the cities (or counties) where the participant was raised; *Outside* indicates the number of years the participant has lived outside coastal Southern California.

Speaker ID	Gender	Age	Ethnicity	Outside	SoCal Origin
Cal001	F	21	White	0	Los Angeles County
Cal002	F	20	White	0	Long Beach
Cal003	F	19	White	0	San Marcos
Cal004	F	21	Vietnamese	0	Santa Ana, San Diego
Cal006	F	20	Latina	0	Chino, La Jolla
Cal007	M	22	White	0	Sun Valley, Thousand Oaks
Cal008	M	18	Asian	0	Rowland Heights
Cal009	F	21	Mexican-American	0	Garden Grove
Cal010	F	34	Filipino	0	San Diego
Cal011	M	20	Afghan	0	Laguna Niguel
Cal012	M	21	White/Asian	0	Camarillo, Northridge
Cal013	M	18	Mixed	0	Orange County
Cal014	M	18	Filipino	0	Walnut
Cal015	M	31	White/Mexican	0	San Diego
Cal016	F	18	Filipino	0	Lake Elsinore

Table 2.2: Demographic information for South Carolina participants. *SC Origin* indicates the counties in South Carolina where the participant has lived; *Outside* indicates the number of years the participant has lived outside South Carolina.

Speaker ID	Gender	Age	Ethnicity	Outside	SC Origin
SC001	M	30	White	0	Richland
SC002	F	27	White	1	Lexington, Richland
SC003	F	22	White	0	Spartanburg, Richland
SC004	F	27	White	3	Berkeley, Dorchester, Richland
SC005	F	20	White	0	Greenville, Richland, Lexington
SC007	F	50	White	4	Greenville, Spartanburg, Richland
SC008	M	27	White	4	Aiken, Richland
SC009	F	18	White	0	Kershaw, Richland
SC010	F	27	White	2	Kershaw, Richland
SC011	F	22	White/Latina	0	Charleston, Richland

2.3.3 Materials

The wordlist for this experiment is presented in Appendix A. It contains 203 mostly monosyllabic words of English, with a small number of disyllabic words where lexical gaps exist. In disyllabic words, primary stress falls on the target vowel. The vowels included in the wordlist were /i u ɪ ʊ e o ɑ ɔ/, which include the three back vowels (/u o ʊ/) observed in acoustic studies to undergo fronting, their front unround counterparts (/i e ɪ/), and the low back vowels /ɑ/ and /ɔ/, which do not undergo fronting in these varieties. Each vowel appeared in a variety of phonological contexts, with onsets including voiceless labial, coronal, and dorsal stops, as well as the voiceless fricatives /s/, /ʃ/, and /h/. In some cases, voiced onsets were used to overcome lexical gaps. Coda consonants included the voiceless stops /p t k/, as well as lateral /l/. Codas were restricted to voiceless stops in order to avoid the effects of coda voicing on preceding vowel length. /l/ was included because it is known to inhibit vowel fronting, particularly for /u/.[6] Vowel-final (but not vowel-initial) words were also included.

2.3.4 Procedure

Recording for this experiment took place in two locations, in sound-attenuated rooms at the University of South Carolina and at the University of California, San Diego. Identical methods and equipment were used in both locations. Ultrasound data were captured using an Articulate Instruments SonoSpeech Micro ultrasound system with a 20mm radius 2–4MHz transducer. During recording, participants were seated with the ultrasound transducer held in place beneath their chin with a stabilizing headset (Articulate Instruments Ltd. 2008). Sagittal-view video of the speaker's lips was captured at 60 fps using a camera mounted

6. In the following analysis, pre-lateral /u/ is referred to as /ul/. Unless otherwise indicated, the symbol /u/ should therefore be taken to refer to tokens of /u/ not before /l/.

to the ultrasound headset. Audio was captured with an AKG C544 L cardioid headset condenser microphone and recorded at a 48 kHz sample rate and 16-bit sample depth with a Marantz PMD661 Mk2 solid state recorder. In addition, audio was recorded directly to disk in Articulate Assistant Advanced (AAA; Articulate Instruments Ltd. 2012), which was used to synchronize the audio, video, and ultrasound data streams.

Participants were asked to repeat the wordlist described above, with each word spoken in the carrier phrase "say ____ again." Participants produced three repetitions of each phrase, which provided 609 tokens per participant, for a total of 15,225 tokens across all participants.[7] Words were presented in pseudorandom order, so that no two words containing the same vowel appeared in successive order. In addition, the vowels /u o ʊ/ did not appear in sequence with their front unround counterparts, /i e ɪ/, and vice versa. The order of presentation was unique for each participant. Prompts were automatically presented to participants in AAA, with the timing of presentation determined by the participant's normal speech rate during the practice phase of the experiment. The practice phase consisted of 3-5 trials in which participants repeated a set of words containing vowels not otherwise in the wordlist, in this case /æ ɛ ɝ/. In addition, a palate trace was captured by asking participants to hold a bolus of water in their mouth and swallow at the experimenter's instruction. The total duration of this procedure was approximately 35 minutes.

2.3.5 Data Analysis

Acoustic data were analyzed in Praat v6.0.36 (Boersma and Weenink 2017). A TextGrid was automatically created for each recording based on the prompt file exported from AAA. FAVE-align v1.2.2 (Rosenfelder et al. 2015) was used to force-align the phonetic transcription. Target intervals were manually corrected, with vowels considered to begin at the start of periodicity. Vowels were considered to end at the point where fewer than two formants

7. This figure does not take into account data excluded due to intermittent mispronunciations, etc.

were clearly visible, where there was a change in formant structure or complexity of the waveform (in the case of lateral codas), or at the beginning of glottalization. In addition, vowels with a duration of less than 80 ms were excluded.

LPC formant measurements were taken using the Formant object in Praat, with LPC coefficients calculated using the Burg algorithm (Childers 1978; Press et al. 1992). For most speakers, formants were computed with 10 poles, with the maximum formant value set to 5000 Hz for men and 5500 Hz for women. Formant measurements were taken at 25% and 75% of the vowel's duration, as well as at the point of maximum labial articulation. Vowel formant measurements were normalized according to the Nearey1 normalization procedure (Nearey 1978) using the vowels package for R (Kendall and Thomas 2014; R Core Team 2018). The Nearey1 method is a vowel-extrinsic, formant-intrinsic normalization procedure which uses log means to scale the value for each formant with individual scaling factors. This method was chosen because, unlike the Lobanov z-score normalization method, it allows for the normalization of F3, which was considered as a metric for lip rounding. Consideration of the normalized F3 measurements is left for future analyses, however. The Nearey1 method was found by Adank, Smits, and van Hout (2004) to be among the best-performing normalization procedures for minimizing the effects of physiological variation while preserving phonemic categories and sociolinguistically-relevant variation.

Ultrasound data were analyzed in Articulate Assistant Advanced v217.05 (Articulate Instruments Ltd. 2012). Tongue splines were automatically fit to the ultrasound data using the Batch Process function. The search space was defined by manually setting Roof and Minimum Tongue splines for each speaker on the basis of frames containing the palate trace, the point of maximum tongue backness (from tokens containing pre-lateral back vowels), the point of maximum tongue lowering (from tokens containing /ɑ/ or /ɔ/), and the point of maximum tongue root advancement (from tokens containing /i/ and /e/). Automatically splined tongue contours were checked for accuracy and manually corrected when neces-

sary. Still images corresponding to each splined frame were exported along with the tongue spline coordinates, and the frame containing the point of maximum lingual articulation was selected for analysis. Points along the tongue contour with a confidence level of less than 100 were excluded from the dataset, as were points that fell beyond the length of the image of the tongue surface.

Lip video data were analyzed with a purpose-built tool written in Python using PsychoPy (Peirce 2007). The TextGridTools package for Python (Buschmeier and Włodarczak 2013) was used to read the hand-corrected TextGrids and identify the start and end points of the target vowel intervals. FFmpeg (FFmpeg Developers 2018) was then used to extract still frames from the portion of the video corresponding to the vowel, plus the preceding and following 50 milliseconds for context. For each target vowel, the annotator was prompted to scroll through the extracted video frame-by-frame and identify the point of maximum labial articulation. Points were manually placed at the upper and lower edges of the lip aperture, respectively defined as (i) the boundary between the vermilion border and oral mucosa of the upper lip and (ii) the nearest point on the lower lip. A third point was placed at the oral commissure. Vertical lip openness was calculated as the Euclidean distance between the upper and lower points. The degree of horizontal lip spread was then determined by calculating the horizontal distance between the oral commissure and the plane intersecting with the upper and lower lip aperture points. Finally, the degree of lip protrusion was determined by calculating the horizontal distance of the point placed on the lower lip from the posterior edge of the video frame. Although coronal-view video was also recorded, only the sagittal-view video is considered at present.[8]

8. In addition, the present analysis relies on lip protrusion, rather than lip spread or openness, as a metric for lip rounding; additional lip rounding metrics will be incorporated in future analyses.

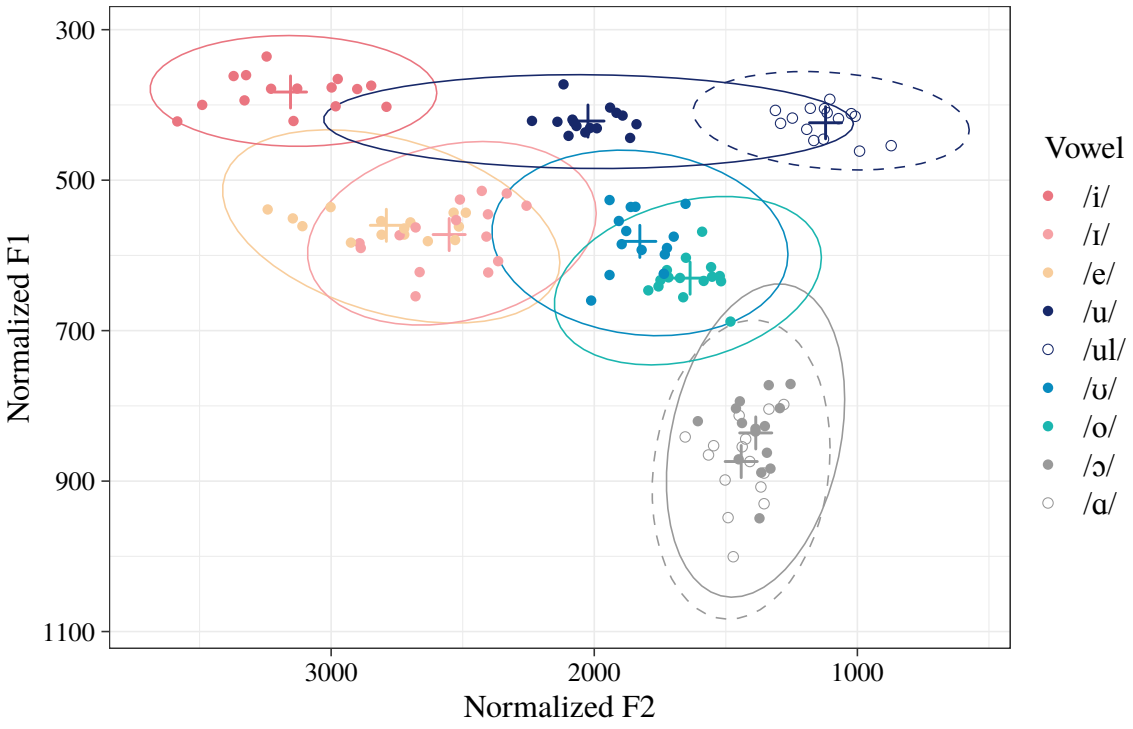


Figure 2.4: Normalized mean formant measurements for Southern California speakers. Measurements were taken at 25% of the vowel's duration. Individual points indicate vowel category means for each speaker; cross marks indicate group means. Ellipses indicate 95% confidence intervals. Pre-lateral tokens other than /ul/ excluded.

2.4 Results for Southern California Speakers

2.4.1 Acoustic Patterns

Normalized mean formant measurements for Southern California speakers are presented in Figure 2.4. This plot includes group and individual mean formant measurements for each of the vowels included in the wordlist, in word-final or pre-obstruent contexts, except for /ul/, which is included as a reference point for backness. First, it is observed that the distributions of /ɑ/ and /ɔ/ exhibit nearly complete overlap, which is expected given that most speakers

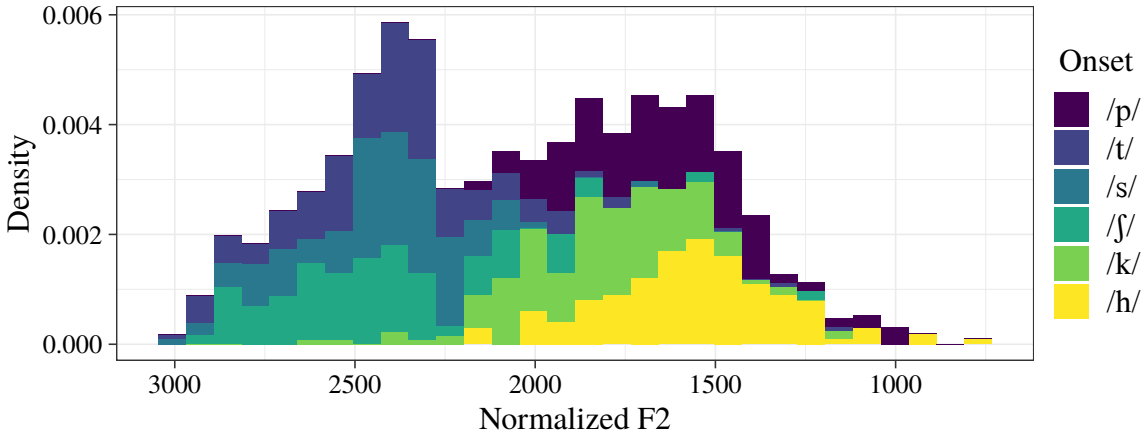


Figure 2.5: Histogram of normalized F2 measurements for /u/ by onset, Southern California speakers. Measurements taken at 25% of vowel's duration.

in the western United States exhibit a merger of these two vowels (Labov, Ash, and Boberg 2006, 61).[9] The most notable observation revealed in this plot is that /u/ exhibits a wide range of values for F2, such that some tokens have an F2 nearly as high as /i/, with values for F2 above 2800 Hz, while other tokens have a relatively low F2, with values below 1300 Hz.[10] The mean F2 for all tokens of /u/ is 2024 Hz, which falls in the central region of the vowel space. /ʊ/ and /o/ also exhibit fronting, with /ʊ/ more advanced than /o/, but less fronted than /u/. The mean F2 for /ʊ/ is 1827 Hz, while the mean F2 for /o/ is 1636 Hz. Both vowels have mean values for F2 greater than those for /ul/ and /ɔ/, but for most speakers, these vowels remain back of the center of the vowel space.

Figure 2.5 displays normalized F2 measurements for /u/, with tokens categorized by preceding consonant. It is observed that the distribution of F2 for /u/ is roughly bimodal,

9. For California speakers, the measurements for these two vowels are henceforth combined into a single vowel category, /ɔ/.

10. These upper and lower values are the 95th and 5th percentiles, respectively, rounded to the nearest 50 Hz.

Table 2.3: Linear mixed effects regression model for F2 of /u/, Southern California speakers.

Predictor	Estimate (Hz)	SE	z value	Pr(>|z|)	
Intercept (Glottal)	1558.4	48.5	32.1	<0.001	***
Onset					
Labial	93.3	53.5	1.7	>0.05	
Dorsal	270.6	53.3	5.1	<0.001	***
Coronal	848.7	46.1	18.4	<0.001	***

with higher values for F2 following coronal onsets and lower values for F2 following non-coronal onsets. Following coronal onsets, including /t/, /s/, and /ʃ/, the mean F2 for /u/ is 2385 Hz. Following non-coronal onsets, the mean F2 is 1811 Hz.

Table 2.3 presents a linear mixed effects regression model for F2 of /u/, with a fixed effect of onset place of articulation and random effects of speaker and word.[11] The model was built using the lme4 package for R, with p-values for each effect from lmerTest (Bates et al. 2015; Kuznetsova, Brockhoff, and Christensen 2017; R Core Team 2018). This model shows that the place of articulation of the onset consonant is a significant predictor of F2 for /u/. The intercept of the model is tokens of /u/ following a glottal onset, which are predicted to have an F2 of 1558 Hz. Following labial onsets, F2 is predicted to be higher, but not significantly so. However, a significant increase in F2 is predicted following velar and coronal onsets. For coronal onsets in particular, the F2 is predicted to be substantially higher, around 2407 Hz, which is close to the actually observed mean F2 for post-coronal /u/ of 2385 Hz. This distribution suggests that the strong fronting of /u/ following coronal onsets is the result of coarticulation from the fronted tongue position required to produce /t/, /s/, and /ʃ/.

11. A model was also built with an additional fixed effect of coda place of articulation, but the fit of that model was not significantly better than the model without this effect.

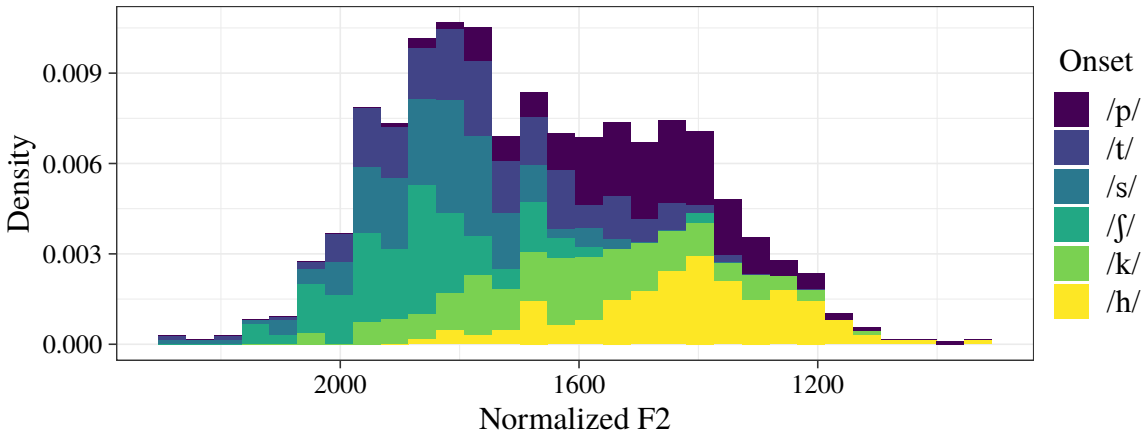


Figure 2.6: Histogram of normalized F2 measurements for /o/ by onset, Southern California speakers. Measurements taken at 25% of vowel's duration.

A similar, but less strongly bimodal, pattern is observed for F2 of /o/, as shown in Figure 2.6. The mean F2 for /o/ is 1795 Hz in post-coronal environments, and 1539 Hz when preceded by labial, velar, or glottal onsets. A linear mixed effects regression model was also built for F2 of /o/, as shown in Table 2.4. It is again observed that F2 for /o/ is predicted to be significantly higher following dorsal and coronal onsets than following labial or glottal onsets. The effect of coronal onsets in particular is much weaker for /o/ than for /u/, as indicated by the much smaller estimate, but these findings nevertheless suggest that fronting is associated with coarticulatory effects from coronal onsets.

2.4.2 Articulatory Patterns

In order to visualize the tongue shapes used to produce these vowels, the tongue splines were analyzed with polar smoothing spline ANOVA. SS ANOVA is described by Gu (2002) and has been used in linguistic research to analyze both ultrasound tongue contour data

Table 2.4: Linear mixed effects regression model for F2 of /o/, Southern California speakers.

Predictor	Estimate (Hz)	SE	z value	Pr(>|z|)	
Intercept (Glottal)	1427.6	55.0	26.0	<0.001	***
Onset					
Labial	69.4	62.3	1.1	>0.05	
Dorsal	189.4	65.2	2.9	<0.001	***
Coronal	383.4	56.2	6.8	<0.001	***

(Davidson 2006; Chen and Lin 2011; De Decker and Nycz 2012; Lee-Kim, Davidson, and Hwang 2013; Lee-Kim, Kawahara, and Lee 2014), and formant measurements (A. Baker 2006; Nycz and De Decker 2006; Fruehwald 2010). The SS ANOVA tongue contours for this experiment were calculated in polar coordinates (following Mielke 2015) in order to avoid distortion of the tongue shape. When calculated in Cartesian coordinates, SS ANOVA makes comparisons along the *x*-axis (i.e., between measurements with the same value for *x*). While this is acceptable for comparing sections of the tongue that are roughly parallel to the *x*-axis, such as the tongue body, it introduces distortions of the tongue shape in regions where the tongue is perpendicular to the *x*-axis, such as the tongue root. This is clearly undesirable for a study of tongue fronting, which to a large extent involves differences in tongue root position.

Figures 2.7 and 2.8 present SS ANOVA tongue contours for two speakers from Southern California. These plots show best-fit smoothing splines for each vowel category, along with 99% Bayesian confidence intervals. Splines were rotated to orient the top of the plot with the top of the speaker's head.[12] Where the confidence intervals for two contours overlap,

12. Because the occlusal plane was not directly imaged for these speakers, splines were rotated on the basis of the lip video. The ultrasound probe and camera are fixed with respect to one another, making it possible to visually determine the approximate rotation of the camera relative

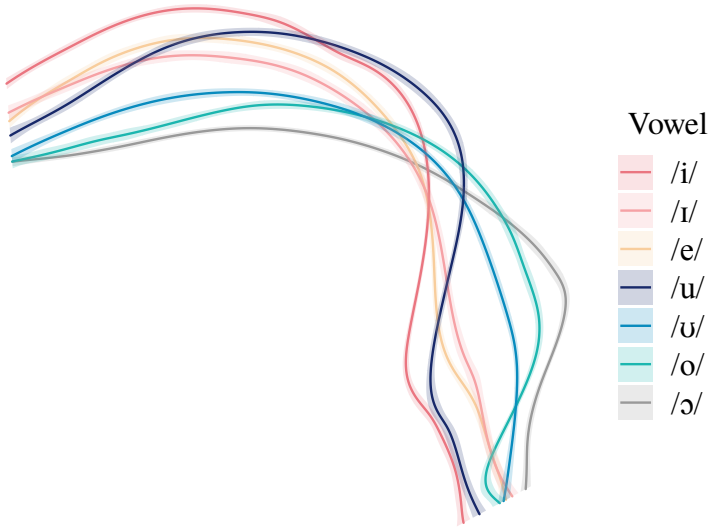


Figure 2.7: Smoothing spline estimates for Cal007, all vowels. Tongue front is to the left, tongue root is to the right. Shading indicates 99% confidence interval. Splines rotated 10.6° clockwise to orient the occlusal plane approximately to horizontal.

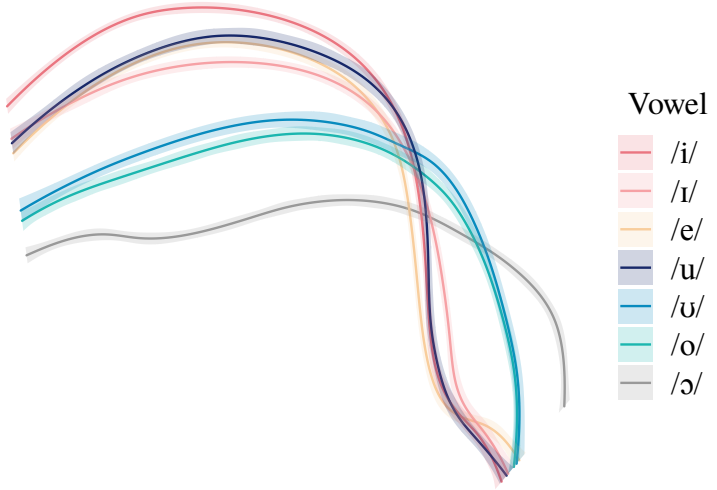


Figure 2.8: Smoothing spline estimates for Cal008, all vowels. Tongue front is to the left, tongue root is to the right. Shading indicates 99% confidence interval. Splines rotated 11.7° clockwise to orient the occlusal plane approximately to horizontal.

the difference between the contours is not statistically significant. For both speakers, the tongue position for /e/ is higher than that of /ɪ/, which is unexpected based on the canonical descriptions of these vowels, but such reversals of tongue height have been observed in a number of previous articulatory studies (Ladefoged et al. 1972; Noiray, Iskarous, and Whalen 2014). The tongue position for /ɔ/ follows the expected pattern: for both speakers, it is the vowel with the lowest tongue height and greatest degree of pharyngeal constriction. The tongue positions for the vowels /ʊ/ and /o/ fall in between those for /u/ and /ɔ/ in terms of both tongue height and backing. For Cal007, /o/ is significantly lower and backer than /ʊ/, while for Cal008, /o/ is significantly lower than /ʊ/, but not significantly backer. Finally, Figure 2.7 reveals that Cal007 produces /u/ with with a tongue position that is significantly backer than the tongue position for /i/, which (as expected) is the frontmost vowel. Cal008, on the other hand, exhibits an extremely fronted tongue position for /u/, which is not significantly different from /i/ for most of the tongue root and some of the dorsum, as seen in the right side of Figure 2.8. Thus, the results from these two speakers suggests that there is a range of variability in tongue position for /u/.

However, one shortcoming of SS ANOVA for the analysis of ultrasound data is that while SS ANOVA models offer a holistic, qualitative interpretation of the data, they do not provide a quantitative value that can (easily) be subjected to further statistical analysis. While SS ANOVA is useful for visualizing the overall shape and position of the tongue for a given sound, and for determining whether significant differences exist between the tongue shapes for two sounds, it provides no information with respect to *how* different two tongue shapes are, only *whether* they are different. Whereas Figure 2.8 shows that Cal008 does not produce /i/ and /u/ with a significant difference in tongue root position, there is no way to

to the speaker's head and use this angle to correct the orientation of the ultrasound splines. While more sophisticated means are necessary for quantitative analysis of (for example) tongue height, this method is deemed sufficient for illustrative purposes. Moreover, rotating the splines does not affect their shape or their associated confidence intervals, so any interpretations made on the basis of the confidence intervals remain valid.

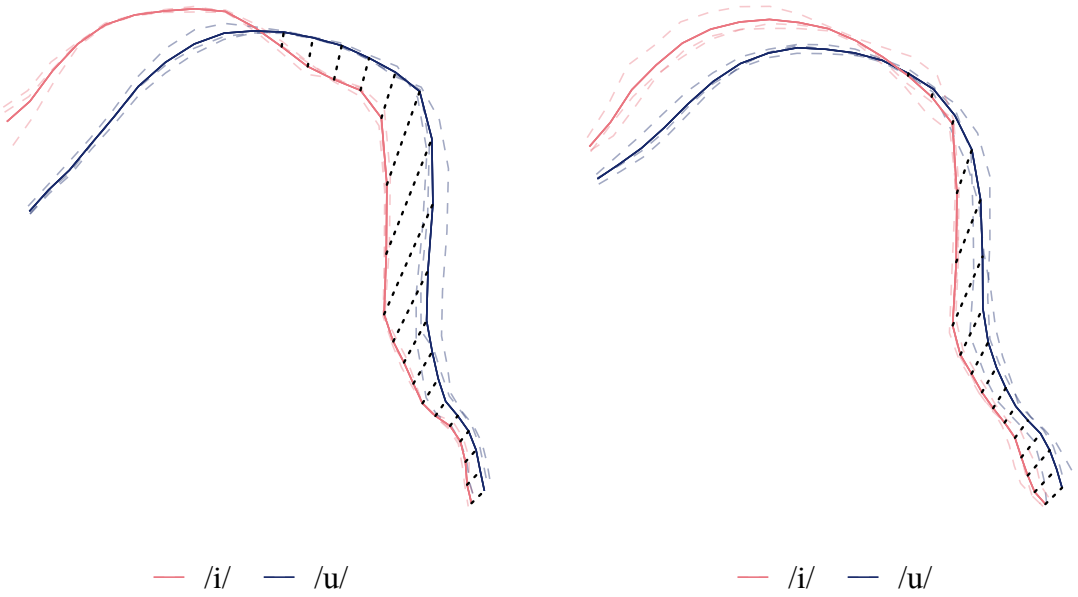
assess the degree of fronting or advancement of /u/ for speakers like Cal007, who constitute the majority of the speakers in this study.

In order to quantify the degree of tongue fronting across speakers, a measure of "summed radial difference" (RD-Σ) between the tongue shapes for /i/ and /u o ʊ/ was calculated. Ultrasound tongue spline data were exported from AAA in polar coordinates, with each spline comprising forty-two points defined in terms of polar angle (θ) and distance (radius) from the center of the ultrasound transducer (in millimeters). For each speaker, the mean tongue contour for all three repetitions of each word was determined by calculating the mean radius along each polar angle. Then, the mean tongue contours for minimal pairs containing the vowels /i/ and /u/, /i/ and /o/, and /i/ and /ʊ/ were compared.[13] The RD-Σ for a given minimal pair (i.e., phonological environment) is the sum of the difference in radius between the two mean tongue contours, for all polar angles where the radius for /u/ (or /o/ or /ʊ/) is greater than that of /i/.[14] In order to make useful interspeaker comparisons, RD-Σ measurements were calculated for all vowels other than /i/ and z-score normalized to capture the full range of tongue positions for each speaker.

The RD-Σ metric is illustrated in Figure 2.9, which shows tongue spline comparisons for the vowels /u/ and /o/ in two phonological environments, as well as (non-normalized) RD-Σ measures. This figure reveals that the radial difference between /u/ and /i/ is smaller in the t_# environment (Figure 2.9b) than for the h_# environment (Figure 2.9a), indicating that the tongue position for /u/ is fronter after /t/ than after /h/. Likewise, the radial difference between /o/ and /i/ is smaller in the word *toe* (Figure 2.9d) than in the word *hoe* (Figure 2.9c). RD-Σ is inversely proportional to tongue fronting, so more fronted tongue positions exhibit a lower RD-Σ.

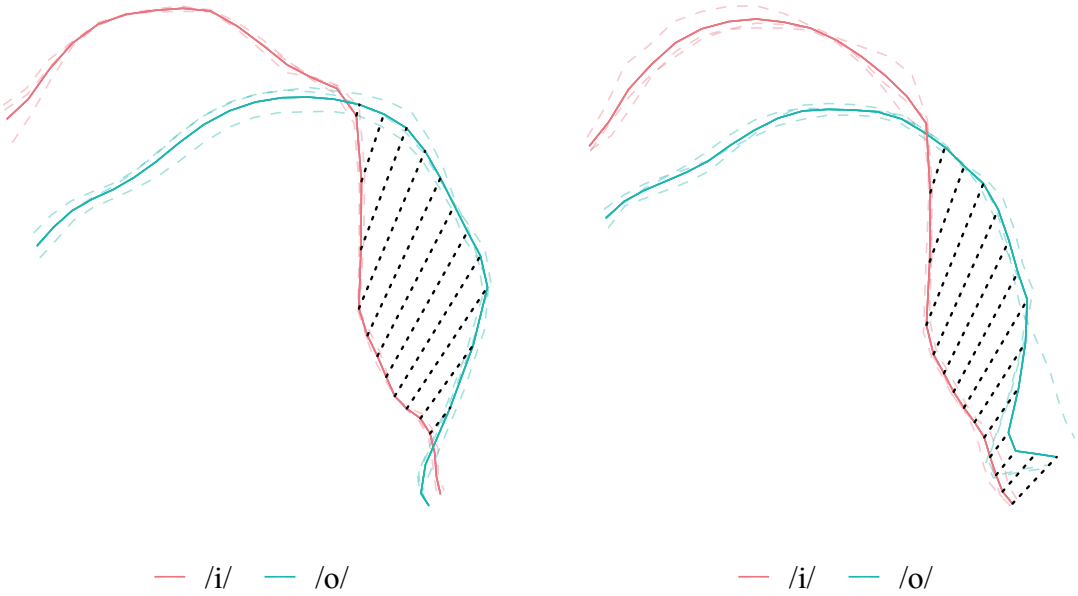
13. Due to lexical gaps or taboo words, some comparisons were made between near-minimal pairs that differed in onset voicing, e.g., *peep* vs. *boop*.

14. Because the RD-Σ method calculates the sum of all radial differences where the radius for the back vowel is greater than that of /i/, any vowels produced with a tongue position fronter than that of /i/ will receive a radial difference measure of 0. This is a point to address in future research.



(a) *he* vs. *who*, RD-Σ = 89.03 mm

(b) *tea* vs. *too*, RD-Σ = 57.6 mm



(c) *he* vs. *hoe*, RD-Σ = 153.05 mm

(d) *tea* vs. *toe*, RD-Σ = 138.74 mm

Figure 2.9: Illustration of the summed radial difference (RD-Σ) metric for determining the degree of tongue fronting for the vowels /u o ʊ/. Dashed lines represent individual tokens, solid lines indicate mean tongue contours for each phonological environment. RD-Σ is the summed length of the black dotted lines. Data from Cal007.

This method was inspired by a metric used by Scobbie and Cleland (2017) for the analysis of velar fronting among children with disordered speech. Scobbie and Cleland present two methods for analyzing the degree of velarity of consonant gestures. The first is the area-based "dorsal crescent" measurement, which gives the area between two tongue contours calculated as the sum of the area of annular sectors centered around each polar angle. The inner radius of each sector is defined by the more coronal tongue contour, while the outer radius is defined by the more velar contour. The second method they propose is the "radial difference" method, which takes the maximal radial distance between the two tongue contours along a single polar angle. The RD-Σ metric employed here falls between these two methods. Like the radial difference method, it is a radius-based, rather than area-based, measure but like the dorsal crescent method, it incorporates differences between the two tongue contours along multiple polar angles, rather than the single greatest distance between the two curves. An advantage of all three measurements is that these methods are immune to interspeaker differences in the angle of the ultrasound probe relative to the speaker's head. This is not the case for methods that rely on identifying tongue position maxima in the *x*, *y* space, such as the highest point of the tongue, which can vary depending on the position of the probe.[15]

Figure 2.10 presents the RD-Σ metric for each word containing /u/, plotted against the mean F2 for that word. This plot reveals that F2 and RD-Σ are negatively correlated, indicating that the closer the tongue position for /u/ is to that of /i/ (resulting in a lower RD-Σ), the higher the value of F2. This correlation demonstrates that the fronting of /u/ is (at least partly) the result of tongue fronting.

In order to determine whether fronted tokens of /u/ and /o/ exhibit lip unrounding in addition to tongue fronting, the degree of lip protrusion for these vowels was also analyzed.

15. This issue can be overcome by imaging the occlusal plane, which can then be used to rotate the tongue splines with the occlusal plane oriented horizontally. However, the occlusal plane was not imaged for the speakers in this study.

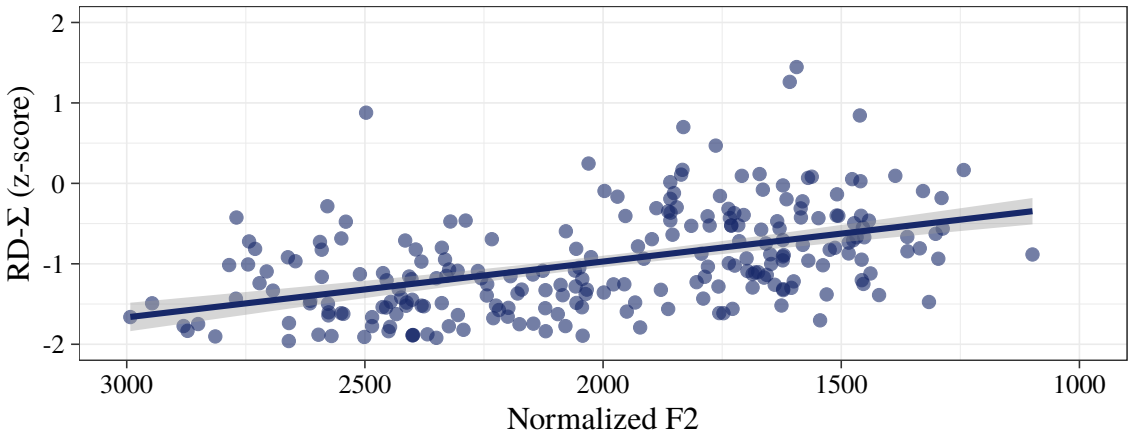


Figure 2.10: Relationship of F2 to tongue frontedness for /u/, Southern California speakers.

Figure 2.11 presents z-score normalized lip protrusion measurements for the two speakers whose SS ANOVA contours are shown in Figures 2.7 and 2.8. This figure, which is representative of the dataset, shows that /u/, /o/, and /ʊ/ exhibit a higher degree of lower lip protrusion than /i/, /e/, and /ɪ/. A one-way ANOVA run for each speaker shows that, for these two speakers, lip protrusion differs significantly between vowel classes (Cal007: $F(7,494) = 221.3$, $p < 0.001$; Cal008: $F(7,475) = 128.6$, $p < 0.001$). A Tukey post hoc test reveals that the difference in lip protrusion for the vowel pairs /i/-/u/, /e/-/o/, and /ɪ/-/ʊ/ is significant at the $p < 0.001$ level. For these two speakers, then, the back vowels appear to have retained at least some degree of lip rounding as they have undergone fronting.

In order to assess whether /u/ or /o/ have undergone unrounding for any of the Southern California speakers, the measurements for F2 were fit against the normalized lip protrusion measurements. If the fronting of /u/ and /o/ is due in part to an unrounding of the lips, the degree of lip protrusion should be lower for tokens with high values of F2 than for tokens

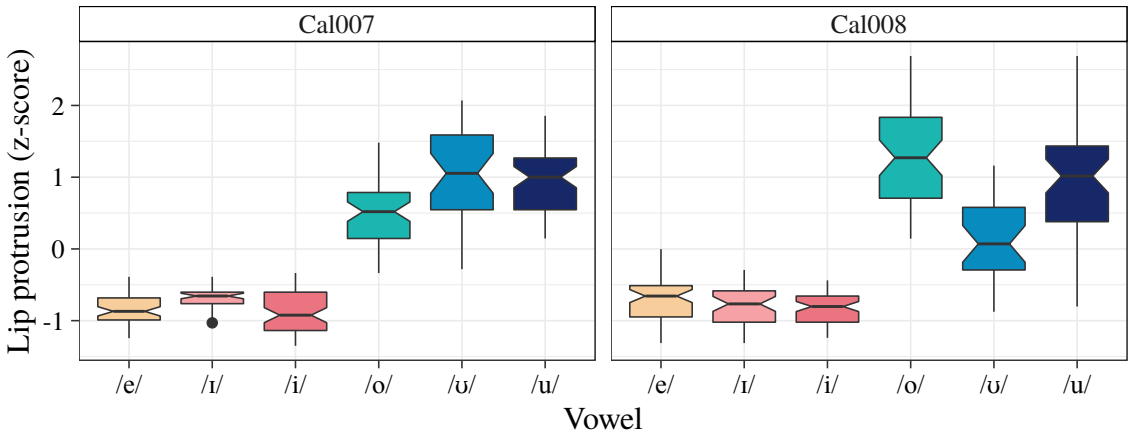


Figure 2.11: Normalized lower lip protrusion in normal speech. Larger values indicate increased lip rounding.

with low values for F2. The relationship between F2 and lip protrusion among Southern California speakers is shown in Figure 2.12. As with the RD-Σ plot in Figure 2.10, this plot contains mean values across the three repetitions of each word in the dataset. F2 measurements in this plot were taken at the same point in the vowel's duration as the measurement for lip rounding, which was measured at the point of maximum labial articulation. Unlike the negative correlation between F2 and RD-Σ, there is no such correlation between lip protrusion and F2. In other words, /u/ exhibits a comparable degree of lip rounding across all tokens, regardless of the value of F2. This suggests that tokens that are acoustically more front are not less round than tokens that are acoustically more back.

The relationship between RD-Σ and lower lip protrusion with F2 for /o/ are presented in Figure 2.13. As for /u/, an increase in F2 for /o/ exhibits a stronger negative correlation with RD-Σ than with lower lip protrusion. Taken together, these results suggest that the fronting

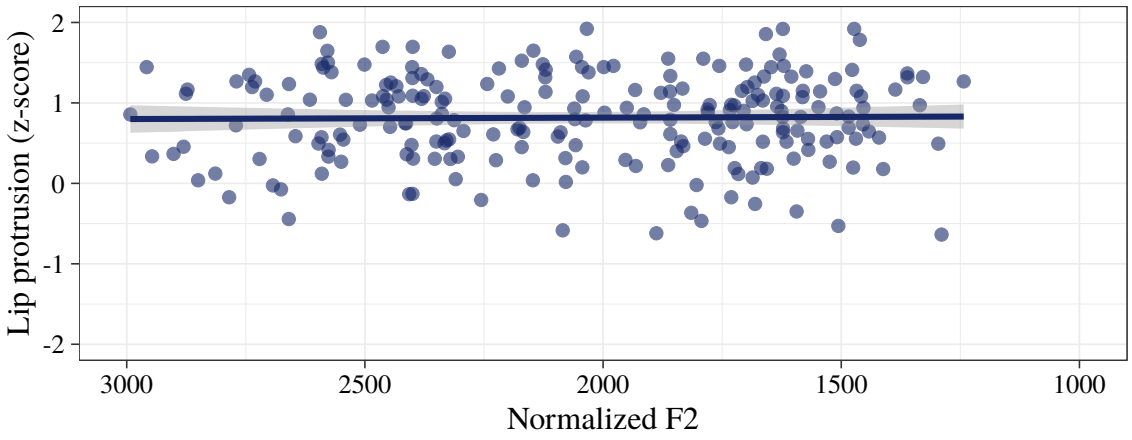


Figure 2.12: Relationship of F2 to lip protrusion for /u/, Southern California speakers.

of /u/ and /o/ in Southern California is achieved primarily by tongue fronting rather than by unrounding of the lips.

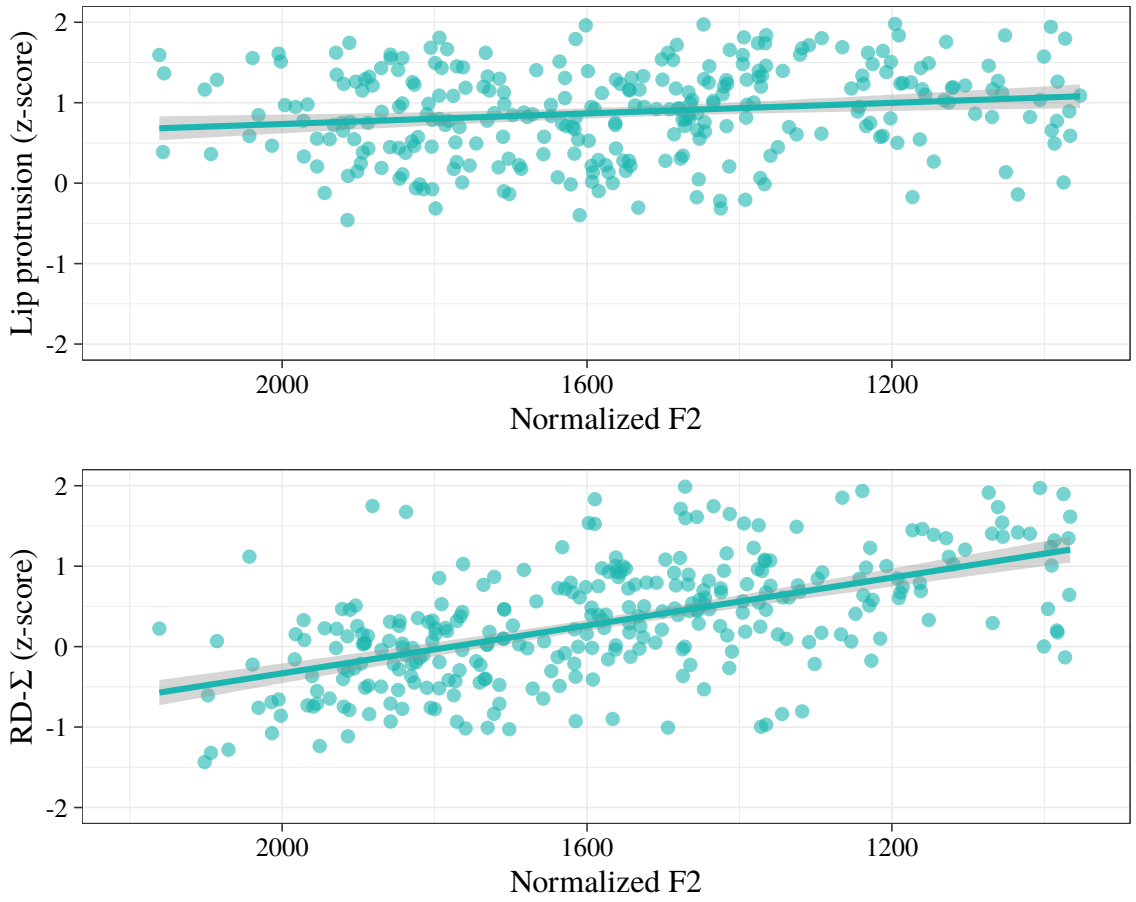


Figure 2.13: Relationship of F2 to lip protrusion (upper) and tongue fronting (lower) for /o/, Southern California speakers.

2.5 Results for South Carolina Speakers

As noted in Section 2.1, the processes of back vowel fronting in Southern California and South Carolina have been observed in acoustic studies to differ, both in the degree to which the back vowels are fronted, as well as in the phonological conditioning of fronting. While the previous section showed that Southern California speakers exhibit strong fronting of /u/ following coronal onsets, previous acoustic studies have shown that /u/ fronting in the South is relatively strong in all phonological contexts, with some speakers showing fronting of /u/ even before laterals. In addition, advanced fronting of /o/ has also been found among speakers from the South, in contrast to many other regions of North America, including California. These differences raise the possibility (but do not necessitate) that these two dialects may also differ in how the fronting of /u/ and /o/ is achieved in terms of articulatory strategy. This section presents both acoustic and articulatory results for ten from South Carolina.

2.5.1 Acoustic Patterns

Figure 2.14 presents normalized mean formant measurements for speakers from South Carolina. As in the vowel chart for Southern California speakers, this chart includes group and individual mean formant values for tokens in pre-obstruent or vowel-final environments. Unlike the speakers from Southern California, the speakers in South Carolina do exhibit a contrast between /ɑ/ and /ɔ/. Two one-way ANOVAs show that F1 ($F(8,4543) = 4472$, $p < 0.001$) and F2 ($F(8,4543) = 2746$, $p < 0.001$) differ significantly between vowels, with Tukey post hoc tests showing a significant difference ($p < 0.001$) between /ɑ/ and /ɔ/ along both dimensions. Like the speakers from Southern California, the speakers from South Carolina produce /u/ with a wide range of values for F2. Tokens at the low end of the range have values near 1400 Hz, while tokens at the high end exhibit values around 2700 Hz. The mean F2 for /u/ is 2073 Hz, which is similar to that observed for California speakers (2024 Hz).

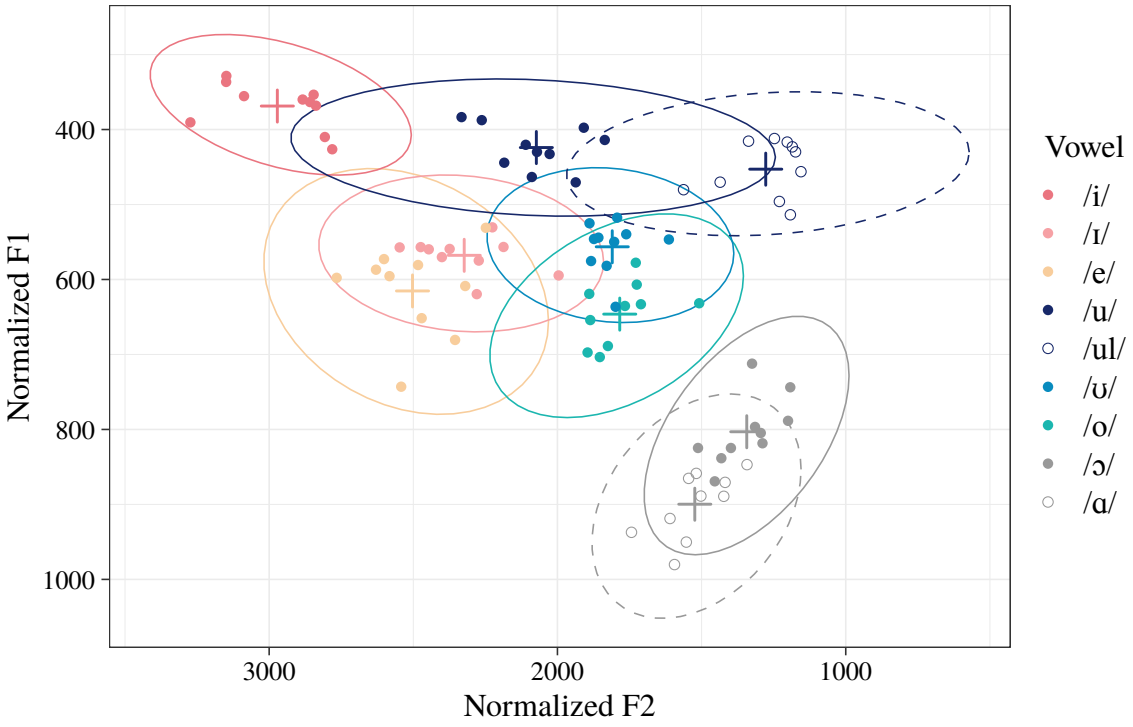


Figure 2.14: Normalized mean formant measurements for South Carolina speakers. Measurements taken at 25% of the vowel's duration. Individual points indicate vowel category means for each speaker; cross marks indicate group means. Ellipses indicate 95% confidence intervals. Pre-lateral tokens other than /ul/ excluded.

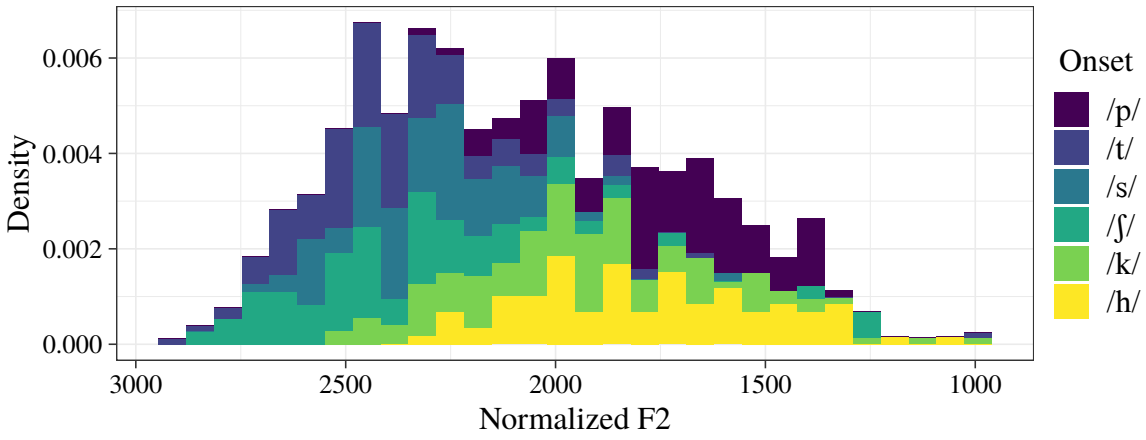


Figure 2.15: Histogram of normalized F2 measurements for /u/ by onset, South Carolina speakers. Measurements taken at 25% of vowel's duration.

In addition, /ʊ/ and /o/ are both fronted for South Carolinians. For /ʊ/, the mean F2 is 1810 Hz, which is comparable to that of the California speakers. /o/ exhibits a greater degree of fronting in South Carolina, such that the mean F2 (1784 Hz) is not significantly lower than that for /ʊ/ ($p = 0.317$). This stands in contrast to the fronting of /o/ among Californians, which lags behind the fronting of /ʊ/.

Figure 2.15 presents a histogram of F2 measurements for /u/, with tokens organized by preceding consonant. As with /u/ in Southern California, fronting of /u/ is more advanced in post-coronal environments, but the distribution of F2 values is far less bimodal in South Carolina. The mean F2 for post-coronal /u/ is 2309 Hz, while the mean F2 for /u/ in all other environments is 1923 Hz. Notably, a number of post-glottal, post-velar, and post-labial tokens of /u/ exhibit an F2 well past the centerline of the vowel space, with some exceeding 2500 Hz.

Table 2.5: Linear mixed effects regression model for F2 of /u/, South Carolina speakers.

Predictor	Estimate (Hz)	SE	z value	Pr(>|z|)	
Intercept (Glottal)	1785.6	71.5	25.0	<0.001	***
Onset					
Labial	-23.6	72.1	-0.3	>0.05	
Dorsal	161.9	72.0	2.2	<0.001	***
Coronal	566.7	62.0	9.1	<0.001	***

A linear mixed effects regression model for /u/, as produced by South Carolinians, is presented in Table 2.5. As with the models built for Southern California speakers in Tables 2.3 and 2.4, a fixed effect of onset place of articulation was included, as were random effects of speaker and word. The model shown here indicates that place of articulation of the onset consonant exerts a significant effect on the F2 of /u/, but to a lesser extent than observed for speakers from Southern California. For South Carolina speakers, the F2 of post-coronal /u/ is predicted to be 567 Hz greater than the F2 of /u/ following a glottal onset. While this increase is statistically significant, it is substantially lower than the 849 Hz increase that is predicted for Southern Californians.

A similar pattern is observed for /o/, as shown in Figure 2.16. Post-coronal tokens of /o/ are concentrated in the high front region of the vowel space, but post-glottal, post-velar, and post-labial tokens are distributed fairly evenly along the F2 dimension. For post-coronal tokens of /o/, the mean F2 is 1892 Hz, whereas the mean F2 following non-coronal onsets is 1714 Hz. A linear mixed effects model for F2 of /o/ was built in the same manner as the models described above. The results of this model are presented in Table 2.6. Again, the effect of a labial onset is not significant, but both dorsal and coronal onsets significantly increase the F2 of /o/.

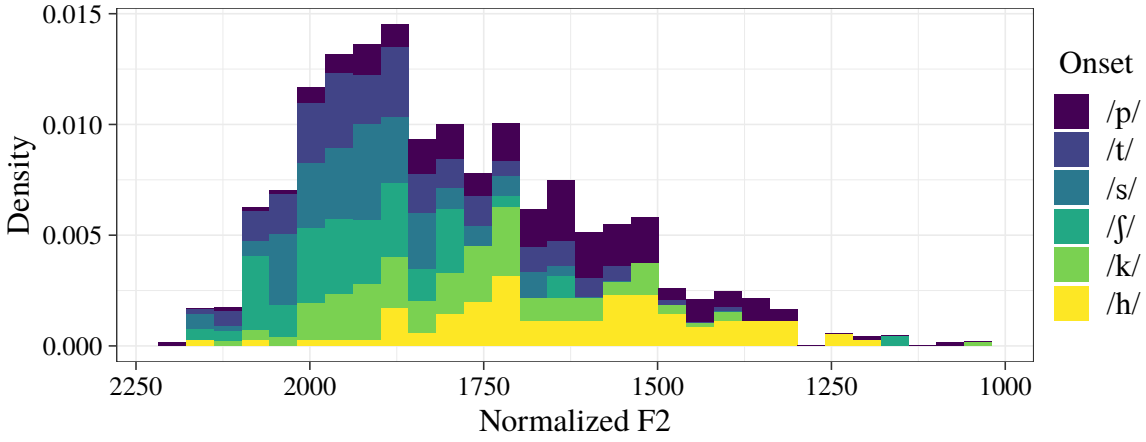


Figure 2.16: Histogram of normalized F2 measurements for /o/ by onset, South Carolina speakers. Measurements taken at 25% of vowel's duration.

Table 2.6: Linear mixed effects regression model for F2 of /o/, South Carolina speakers.

Predictor	Estimate (Hz)	SE	z value	Pr(>|z|)	
Intercept (Glottal)	1625.6	44.3	36.7	<0.001	***
Onset					
Labial	43.3	33.7	1.3	>0.05	
Dorsal	154.7	35.2	4.4	<0.001	***
Coronal	260.3	30.5	8.5	<0.001	***

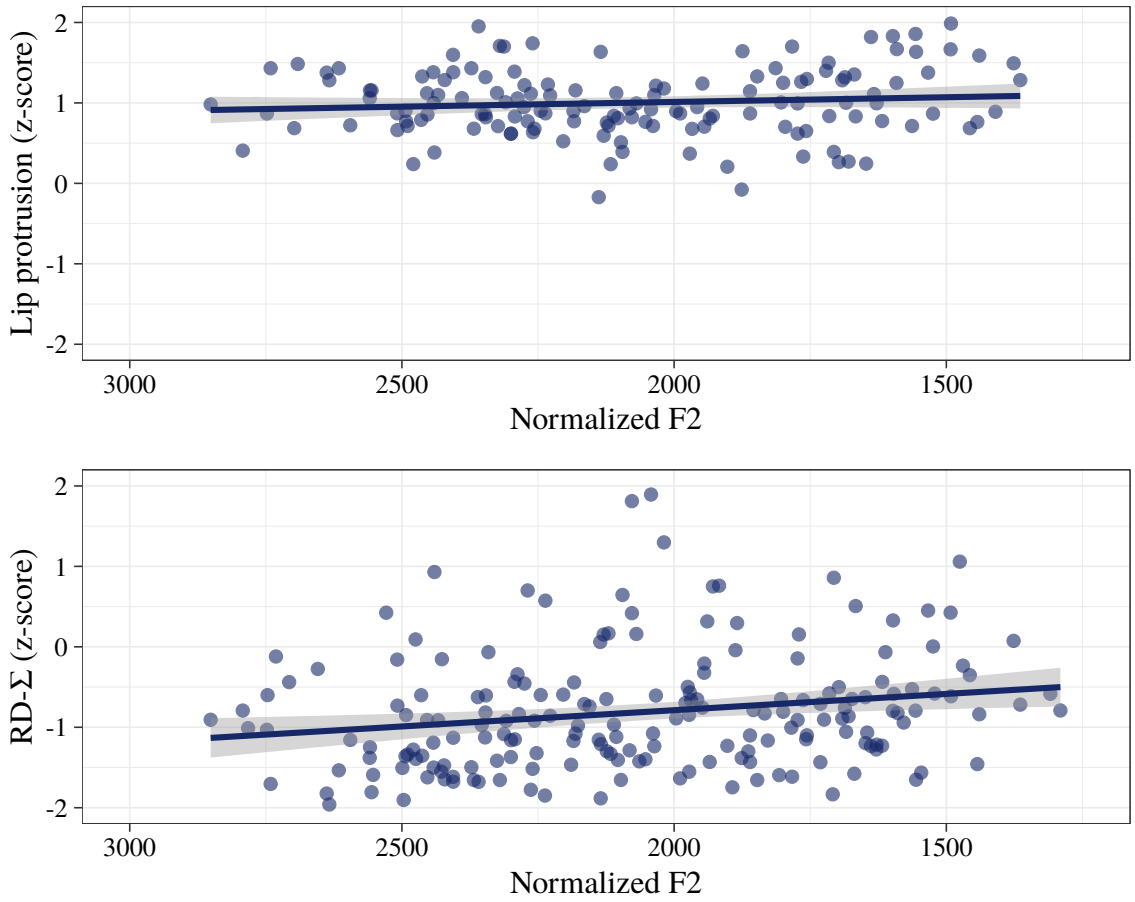


Figure 2.17: Relationship of F2 to lip protrusion (upper) and tongue fronting (lower) for /u/, South Carolina speakers.

2.5.2 Articulatory Patterns

Figure 2.17 presents the relationship between RD-Σ and lower lip protrusion with F2 for /u/, as produced by speakers from South Carolina. As with the speakers from Southern California, RD-Σ and F2 are negatively correlated, with fronter tokens exhibiting a higher F2 than backer tokens. However, it is observed that F2 does not change as a function of

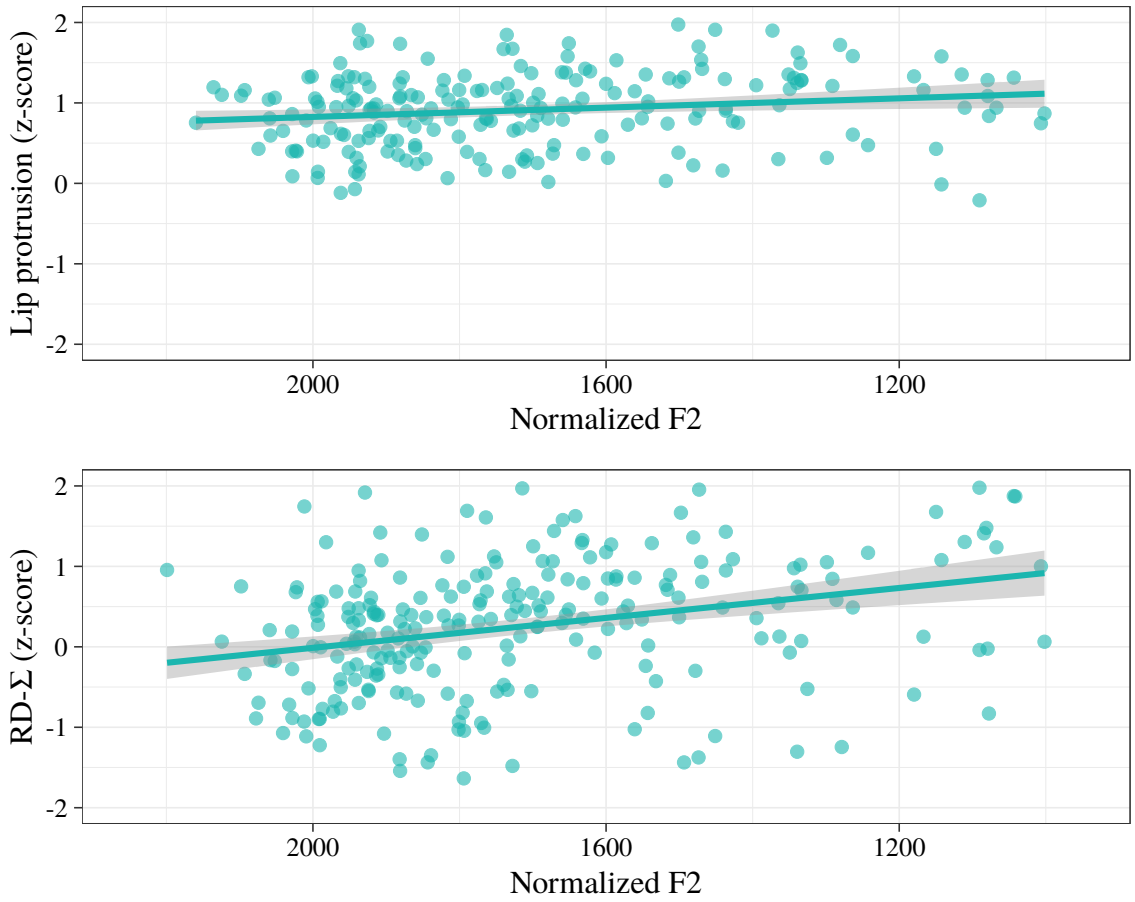


Figure 2.18: Relationship of F2 to lip protrusion (upper) and tongue fronting (lower) for /o/, South Carolina speakers.

lower lip protrusion; tokens with high and low values for F2 exhibit similar degrees of lip rounding, suggesting that tokens with a higher F2 do not exhibit lip unrounding.

A similar pattern is observed in Figure 2.18, which shows that a higher F2 for /o/ is more strongly associated with a fronter tongue position than with unrounded lips.

2.6 Chapter Summary

This chapter has shown that the fronting of the back vowels /u/ and /o/ in these two varieties of American English is achieved primarily by fronting the tongue, rather than by unrounding the lips. While /u/ in particular exhibits a wide range of values for F2, such that some tokens are close to /i/ and others remain in the high back region of the vowel space, the articulatory data showed that acoustically more front tokens exhibit the same degree of lip rounding as acoustically more back tokens. In contrast, acoustically more front tokens show a smaller radial difference measure than acoustically more back tokens, indicating that the tongue position for fronted tokens is closer to the high front tongue position for /i/ than to a back tongue position. Thus, the fronting of /u/ and /o/ is shown to be the result of tongue fronting, and the assertion that these vowels have become unround is not supported. This result is similar to the findings of Harrington, Kleber, and Reubold (2011), who show that /u/ in SSBE exhibits a high front tongue position, but remains round, as confirmed through articulatory and audiovisual perceptual data. Likewise, Scobbie, Lawson, and Stuart-Smith (2012) showed that fronted /u/ in Scottish English is a front or central round vowel. Additional work is needed, however, to determine whether vowel fronting strategies between these dialects differ in other respects, such as in the magnitude or timing of the tongue and lip gestures. Lawson, Stuart-Smith, and Mills (2017) have shown that the fronting of Scottish English /u/ differs from Anglo varieties in the height of the tongue, and suggest that these differences are due to the fronting of /u/ in Scotland being a much older change than the fronting of /u/ in England.

As noted at the start of this chapter, both strategies of tongue fronting and lip unrounding are predicted to be possible in principle, given that both articulatory strategies shorten the front cavity of the vocal tract, thereby increasing the frequency of F2. There are several potential factors that may explain why unrounding is not observed. First, particularly for

speakers from Southern California, the fronting of /u/ and /o/ is strongly associated with coronal onsets. The distribution of F2 for /u/ exhibits a three-way split, with a high F2 following coronals, a low F2 before laterals, and a centralized F2 following labials and velars. Because the coarticulatory source of vowel fronting remains relatively transparent in the acoustic signal, there appears to be no real pressure for /u/ or /o/ to undergo unrounding. It was argued above that a strategy of only unrounding /u/ following coronal onsets, but retaining rounding following labial and dorsal onsets, would be unexpected. Because the coronal consonant would still exert a strong coarticulatory force on the position of the tongue, it would be more difficult to produce [tɯ] or [tɨ] than to produce [ty]. Note, for instance that /tɯ/ sequences are prohibited in Japanese, and are resolved by *s*-epenthesis: [tsɯ]. However, even for speakers from South Carolina, where the fronting of /u/ and /o/ is less strongly associated with coronal onsets, the acoustic fronting of these vowels appears to be the result of tongue fronting, rather than lip unrounding.

While it is not possible (or desirable) to rule out a coarticulatory explanation, another contributing factor for this tendency may be the organization of vowel systems around both auditory and visual/articulatory factors. By retaining the lip rounding gesture for /u/ as it undergoes acoustic fronting, perceptual contrast with /i/ and /ɪ/ is maintained to a greater extent than if /u/ were unrounded to [ɨ] or [ɯ]. In the latter case, /u/ would become similar to /i/ both auditorily as well as visually. If /u/ is realized as [y], on the other hand, it is auditorily similar to /i/ but visually dissimilar. The latter hypothesis, that vowel systems (and phonological systems more generally) are optimized for visual perceptibility, will be explored in more detail in the following chapters. Chapter 6 provides discussion of these results as they relate to theories of sound change.

Chapter 3

Articulatory Strategies for Production of the cot-caught Contrast

In Chapter 2, it was demonstrated that speakers of American English achieve /u/-fronting primarily through tongue fronting, not lip-unrounding, such that /u/ is produced as a high front or high central round vowel. Although it is possible, in principle, to achieve /u/-fronting through either unrounding of the lips or fronting of the tongue, it was hypothesized that the strategy of lip unrounding is less preferred because it results in a less-perceptible contrast between /i/ and /u/ due to the lack of visual speech cues. However, given that the fronting of /u/ is closely tied to its coarticulatory source, such that it is more fronted after coronal onsets, there remain alternate explanations as to why fronted /u/ and /o/ have retained their rounding. Thus, it is worthwhile to investigate a case of unconditioned back vowel fronting, such as the characteristic fronting of /ɑ/ and /ɔ/ in the Northern Cities Shift. The articulatory study presented in this chapter shows that for most speakers, /ɔ/ has retained its rounding as it has undergone fronting. For some of these speakers, /ɑ/ and /ɔ/ are produced with identical tongue positions and are distinguished through lip rounding alone. However, the strength of the contrast between /ɑ/ and /ɔ/ varies widely between speakers, such that younger speakers are less likely to exhibit the contrast than older speakers. The implications of this finding for ongoing sound change in Chicago, and in the Inland North more broadly, are discussed.

3.1 The Northern Cities Shift

The Northern Cities Shift (NCS) arose in the late-nineteenth and mid-twentieth centuries in cities along the Great Lakes shoreline (the Inland North), including Detroit, Chicago, Buf-

falo, Cleveland, and Rochester (Labov, Ash, and Boberg 2006). The shift was first reported by Fasold (1969) in a manuscript on the speech of Detroiters, and has since spread from urban to rural areas throughout the region, as demonstrated by Gordon (1997) and Ito (1999). As a chain shift, the NCS describes the coordinated movement of several vowels, a diagram of which is presented in Figure 3.1. In its earliest stage, that described by Fasold (1969), the NCS involves the movement of /æ/ and /ɑ/. Labov, Ash, and Boberg (2006) argue that the shift originated with the raising of /æ/, which (in many cases) exhibits an F1 as low or lower than for /ɪ/ (Labov 1994). Under their proposal, the raising of /æ/ creates an opening in the vowel space, which /ɑ/ moves forward to fill. Among Inland North speakers, Labov, Ash, and Boberg (2006) find that /ɑ/ exhibits a mean F2 of greater than 1450 Hz, far higher than Peterson and Barney's (1952) finding of a mean F2 for /ɑ/ of 1220 Hz for women and 1090 Hz for men. Thomas (2001) and McCarthy (2010), on the other hand, suggest that the NCS may have begun with the fronting of /ɑ/, rather than with the raising of /æ/. McCarthy (2010) finds support for this analysis in the speech of Chicagoans born between 1891 and 1919; the oldest of these speakers show evidence of /ɑ/ fronting, but only the youngest speakers, born after 1910, show advanced raising of /æ/. In either scenario, the movement of /æ/ and /ɑ/ is followed by the lowering and fronting of /ɔ/, which adopts the former position of /ɑ/. Later stages of the vowel shift involve the movement of several additional vowels, including /ʌ/, /ɛ/, and /ɪ/, the analysis of which forms much of the basis for Eckert's (1989) landmark ethnographic study of Metro Detroit high schoolers. The changes of interest for the present study, however, are the fronting of /ɑ/ and the fronting and lowering of /ɔ/.

Like the fronting of the high back vowels analyzed in Chapter 2, descriptions of the NCS have been based almost entirely on acoustic data,[1] such that the fronting of /ɑ/ and /ɔ/

1. One notable exception is a study by Plichta (2005), who used nasal/oral airflow measurement to investigate the effect of nasalization on /æ/-raising in the NCS. He suggests that raised /æ/ may be an artifact of the high degree of nasal airflow found for Northern Cities speakers in both nasal and non-nasal environments.

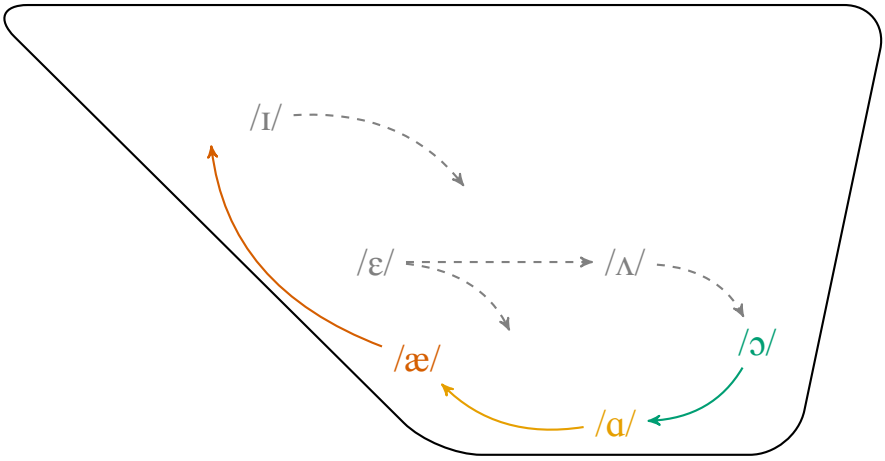


Figure 3.1: Schematic diagram of the Northern Cities Shift. Solid line indicates early stages, dashed line indicates later stages. Adapted from Labov, Ash, and Boberg 2006, p. 190.

is described as an increase in the value of F2. However, an observed increase in F2 can be the result of any gesture that shortens the front cavity of vocal tract, including both tongue fronting and lip unrounding. Thus, it is difficult to determine how this shift is articulatorily achieved based on acoustic evidence alone. Given that /ɑ/ begins as an unround vowel, it is reasonable to assume that its fronting is achieved entirely through tongue fronting. For articulatory change in /ɔ/, however, three possibilities exist. First, the tongue position for /ɔ/ may move forward, approaching that of /ɑ/, while the lips remain round. A second possibility is that /ɔ/ becomes unround with no change in tongue position. Third, these strategies may be combined such that speakers produce fronted /ɔ/ with some degree of lip unrounding and some fronting of the tongue. There is a fourth possibility of merger, which has been reported for speakers in the Inland North only recently.

To investigate this question, Majors and Gordon (2008) used video recording to perform an analysis of lip unrounding in two speakers from St. Louis, where the NCS is in effect to some extent. Majors and Gordon find that /ɔ/ is fronted while retaining its rounding, suggesting that /ɔ/-fronting and lowering in the NCS can be accomplished through a repositioning of the tongue alone. However, because video analysis only allows for measurement of labial articulation, their study is unable to reveal the behavior of the tongue. In addition, St. Louis is the least consistent of the Inland North cities in terms of the number of NCS-related changes and the number of speakers exhibiting the shift (Labov, Ash, and Boberg 2006), so the patterns observed in St. Louis may differ from those found in more typical Inland North cities, such as Chicago or Detroit.

Using a combination of video analysis and ultrasound tongue imaging, Havenhill and Do (2018) find that all three predicted patterns occur among Metro Detroit speakers. While some speakers produce fronted /ɔ/ such that it is distinct from /ɑ/ in both tongue position and lip rounding, others contrast /ɔ/ from /ɑ/ through either tongue position or lip rounding alone. However, these three strategies are not equal with respect to their effects on the acoustic output. Havenhill and Do calculated a Pillai-Bartlett trace ('Pillai score') for each speaker in order to quantify the overlap between /ɑ/ and /ɔ/ in the F1 × F2 space. They found that speakers who distinguish /ɔ/ from /ɑ/ along only one articulatory dimension exhibit a significantly lower Pillai score than do speakers who contrast these vowels along multiple dimensions, indicating that for these speakers, /ɑ/ and /ɔ/ are acoustically more similar than for speakers who produce multiple articulatory distinctions. This result indicates that the use of multiple articulatory gestures to distinguish these vowels may serve to enhance the acoustic contrast. This finding raises the question of how these articulatory and acoustic patterns are perceived and acquired, particularly when /ɑ/ and /ɔ/ are reported to be distinct for most speakers in the Inland North.

3.2 This Experiment

The primary goal of this chapter is to replicate and expand upon the findings of Havenhill and Do (2018), with a larger population of speakers from a previously unstudied (in terms of articulation) city in the Inland North. However, this study also seeks to address several questions (and shortcomings) discussed by Havenhill and Do (2018). First, on the basis of perceptual data, Havenhill and Do (2018) predict that variants of /ɔ/ that involve an unrounding of the lips will be dispreferred. Because their study considered articulatory data from only six speakers, however, it is difficult to make generalizations about the relative frequency of the articulatory patterns observed in their study. They found only one speaker who produced /ɔ/ with no significant difference in lip rounding relative to /ɑ/, but given the small sample size, this result may simply be due to chance. By considering articulatory data from a larger sample of speakers, this study seeks to address the question of whether unround articulatory variants of /ɔ/ are in fact less common than variants that retain their rounding. Given the choice between producing round and unround variants of /ɔ/, both of which may result in similar acoustic output, it is predicted that speakers will prefer variants that rely on lip rounding.

In addition, Havenhill and Do (2018) note that teleological models of sound change and variation, such as those of Lindblom (1990) and Lindblom et al. (1995), predict that speakers will adapt their production patterns for maximal perceptibility when communicating under noisy conditions. If speakers do consider how their speech will be perceived visually, as well as auditorily, it is predicted that speakers who exhibit a contrast between /ɑ/ and /ɔ/ will increase their use of visible lip rounding for the vowel /ɔ/ when attempting to increase the degree of contrast between /ɑ/ and /ɔ/. Such a result would be consistent with the findings of Ménard et al. (2016), who find that sighted (but not blind) speakers increase their use of lip rounding in careful speech. While this prediction does not preclude an increase in the

lingual distinction, the use of non-visible differences in tongue position to the exclusion of a lip rounding enhancement is not predicted to occur.

3.3 Methods

3.3.1 Participants

Sixteen participants (4 men, 12 women) took part in the study, which was conducted at Northwestern University in Evanston, Illinois. Demographic information for each participant is presented in Table 3.1. Participants were natives of the Chicago metropolitan area, having been born and raised in the region through the age of 18. Around half of the participants in the study have lived in the City of Chicago for their entire lives. Seven have lived outside the Chicago area for a period of one or more years, and four have lived at least part of their life in the Chicago suburbs. The age range of participants was 20 to 77 years ($M = 46.5$, $SD = 21$). All participants had self-reported normal hearing and speech as well as normal or corrected-to-normal vision.

Three additional participants (not listed in Table 3.1) also took part in the study but were excluded from the present analysis. One participant, a 71-year-old woman, was excluded because the sagittal-view video did not include the oral commissure, thus preventing measurement of lip spread; this speaker may be included in future analyses that incorporate the coronal-view video that was also recorded. A second participant, a 19-year-old woman, was excluded because the demographic questionnaire revealed that she had lived well outside the Chicago region for more than five years during adolescence; this speaker therefore does not meet the desired degree of autochthony sought for this study. The third participant, a 26-year-old woman, was excluded due to a technical issue that prevented exporting of the ultrasound data.

Table 3.1: Demographic information for Chicago participants. *Areas Lived* indicates the areas of Chicago where the participant has lived; *Outside* indicates the number of years the participant has lived outside the Chicago metropolitan area.

Speaker ID	Gender	Age	Ethnicity	Outside	Areas Lived
CHI001	F	24	white	0	South Side
CHI002	M	56	white	1	Northwest Suburbs
CHI003	F	55	white	0	Far North Side, Northwest Side
CHI005	F	77	white	7	West Side, the Loop, South Side
CHI006	F	66	white	4	Far Southeast Side, Far Southwest Side
CHI008	M	63	white	0	Northwest Side
CHI009	M	20	white	0	South Side
CHI010	M	70	white	0	Northwest Side
CHI011	F	21	white	0	North Shore
CHI012	F	37	hispanic	6	Near West Side, Western Suburbs
CHI013	F	23	white	0	Uptown, Far North Side
CHI015	F	65	white	0	Far North Side
CHI016	F	57	white	4	South Side, Suburbs
CHI017	F	63	white	2	North Side, Far North Side, North Shore
CHI018	F	27	white	2	Outer Suburbs
CHI019	F	20	asian	0	Southwest Side, South Side

3.3.2 Materials

The wordlist for the production experiment consists of 123 mostly monosyllabic English words containing the vowels /i/, /æ/, /u/, /o/, /ɑ/, and /ɔ/. The vowels /i/, /u/, and /o/ were used as reference points for lip spread and rounding, while /æ/ was included as a metric for participation in the NCS.[2] Words containing /ɑ/ and /ɔ/ were the target items and comprise 39% of the wordlist (48 of 123 items). Words were selected such that each vowel appeared in a variety of phonological contexts, including words with coronal, velar, and labial onsets and codas, as well as vowel-initial words. In addition, all of the words used as responses for

2. Labov, Ash, and Boberg (2006) use a mean F1 for /æ/ of less than 700 Hz as one indicator of participation in the NCS, as described in Section 3.4.1.

the perception experiment in Chapter 4 were included in the production task as a control for lexical variation among these words. The complete wordlist is provided in Appendix B.

3.3.3 Procedure

The procedure for the production task generally follows that described in section 2.3.4. Recording took place in a sound-attenuated booth at Northwestern University. Ultrasound data were captured using an Articulate Instruments SonoSpeech Micro ultrasound system with a 20mm radius 2–4MHz transducer. Ultrasound data were recorded at a frame rate of approximately 84 frames per second (fps). Participants were seated with the transducer held in place beneath their chin with a stabilizing headset (Articulate Instruments Ltd. 2008). A sagittal view of the speaker's lips was recorded at 60 fps using a camera mounted to the ultrasound headset and a coronal view was recorded at 120 fps with a Sony RX10-III digital camera. Audio was captured using an AKG C544 L cardioid headset condenser microphone and recorded with an Olympus LS-100 solid state recorder at a 48 kHz sample rate and 16-bit sample depth. Audio, video, and ultrasound data were synchronized in Articulate Assistant Advanced (AAA; Articulate Instruments Ltd. 2012).

Participants were asked to repeat the wordlist two different conditions, in two separate blocks. In the first block, words were elicited in the carrier phrase "say ____ again." Participants produced three successive repetitions of each phrase, for a total of 369 tokens per participant. Prompts were presented in pseudorandom order, such that two words containing the same vowel did not appear in successive order, nor did two words containing the target vowels /ɑ/ and /ɔ/. The prompt list was uniquely randomized for each participant. Prompts were presented to participants on a computer monitor in AAA. Prompts were presented automatically, with the rate of presentation established based on the participant's natural speech rate during the practice phase of the experiment.

In the second block, participants repeated two words containing the target vowel and two words containing a contrasting vowel in the carrier phrase "I said $\underline{target_x}$ and $\underline{target_y}$, not $\underline{contrast_a}$ and $\underline{contrast_b}$," with corrective focus placed on the target word.[3] The items $target_x$ and $contrast_a$ were a minimal or near-minimal pair, as were $target_y$ and $contrast_b$. For example, if the target words were *nod* and *sod* and the contrasting words were *gnawed* and *sawed*, the participant would say: "I said *nod* and *sod*, not *gnawed* and *sawed*." To avoid the effects of phrase-final intonation, only $target_x$ and $contrast_a$ are included in the analysis. Target and contrast items alternated such that each word in the wordlist appeared in all 4 possible positions, for a total of 36 phrases.[4] Because each word was measured in two positions, the careful speech task provided 72 tokens per participant. The total number of tokens measured for both tasks was 7,056, not accounting for tokens excluded due to occasional mispronunciation.

Three practice trials were provided before each block, in order to familiarize participants with the procedure and format of the prompts. Practice words contained the vowels /ɛ ʊ ɪ/ (i.e., vowels not otherwise in the wordlist). A palate trace was captured at the start of recording by asking participants to hold a bolus of water in their mouth and swallow at the experimenter's instruction. The palate trace was used to determine the maximum possible height of the tongue during splining. In addition, the occlusal plane was imaged by asking participants to lightly bite a wide tongue depressor placed horizontally across the participant's tongue and between the upper and lower teeth. The tongue depressor caused the surface of the speaker's tongue to appear as a flat line on the ultrasound image, the angle of which was calculated. The extracted tongue splines were subsequently rotated to orient the occlusal plane horizontally. The total duration of the production task was approximately 25 minutes per participant.

3. The specific instruction given to participants was to "speak clearly and with as much emphasis as possible, as though you are correcting someone who misheard you."

4. The items used as responses in the perception task were not included in the careful speech task.

3.3.4 Data Analysis

Data analysis procedures were similar to those used for the production experiment in Chapter 2. Acoustic data were analyzed in Praat v6.0.36 (Boersma and Weenink 2017). A TextGrid was automatically created for each recording based on the prompt file exported from AAA. FAVE-align v1.2.2 (Rosenfelder et al. 2015) was used to force-align the phonetic transcription. Target intervals were manually corrected, with vowels considered to begin at the start of periodicity. In the case of vowel-initial words, the segment was considered to begin either at the start of periodicity, or at the cessation of glottalization when present. Vowels were considered to end at the point where fewer than two formants were clearly visible, where there was a change in formant structure or complexity of the waveform (in the case of sonorant codas), or at the beginning of glottalization.

LPC formant measurements were taken using the Formant object in Praat, with LPC coefficients calculated using the Burg algorithm (Childers 1978; Press et al. 1992). Except for tokens containing /æ/, measurements were taken at the point of F1 maximum. For /æ/, the point of F2 maximum was used, as suggested by Labov, Ash, and Boberg (2006, 38). Formants were computed with the maximum formant set to 5000 Hz for men and 5500 Hz for women.

Vowel formant measurements were normalized using the *ANAE* log-mean normalization formula (39–40), as implemented in the R package vowels (Kendall and Thomas 2014). The *ANAE* normalization method is a modified form of the Nearey (1978) log-mean normalization method, employing a speaker-extrinsic grand mean (G value). Rather than compute a grand mean for the relatively small dataset used here, the default value of 6.896874 was used, which is the value found for a sample of 345 North American speakers in the Telsur study conducted by Labov, Ash, and Boberg (2006). Although this method of normalization is typically best suited for larger sample sizes, it is beneficial in that it allows for a compar-

ison of the formant values obtained in this study to the values found for Northern Cities speakers in *ANAE*, thereby providing a useful metric for a speaker's degree of participation in the NCS.

Ultrasound data were analyzed in Articulate Assistant Advanced v217.05 (Articulate Instruments Ltd. 2012). Tongue splines were automatically fit to the ultrasound data using the Batch Process function. The search space was defined by manually setting Roof and Minimum Tongue splines for each speaker on the basis of frames containing the palate trace, the point of maximum tongue backness (from tokens containing pre-lateral back vowels), the point of maximum tongue lowering (from tokens containing /ɑ/ or /ɔ/), and the point of maximum tongue root advancement (from tokens containing /i/). Automatically splined tongue contours were checked for accuracy and manually corrected when necessary. Still images corresponding to each splined frame were exported along with the tongue spline coordinates, and the frame containing the point of maximum lingual articulation was selected for analysis. Points along the tongue contour with a confidence level of less than 100 were excluded from the dataset, as were points that fell beyond the length of the image of the tongue surface.

Tongue contours were analyzed using smoothing spline analysis of variance (SS ANOVA), which, as noted in Chapter 2, is a statistical method for determining whether significant differences exist between best-fit smoothing splines for two or more sets of data. Tongue contour data were exported from AAA in polar coordinates, which facilitated fitting of the SS ANOVA models in polar coordinates (following Mielke 2015), as well as rotation of the tongue contours.

Lip video data were analyzed with a purpose-built tool written in Python using PsychoPy (Peirce 2007). The TextGridTools package for Python (Buschmeier and Włodarczak 2013) was used to read the hand-corrected TextGrids and identify the start and end points of the target vowel intervals. FFmpeg (FFmpeg Developers 2018) was then used to extract still

frames from the portion of the video corresponding to the vowel, plus the preceding and following 50 milliseconds for context. For each target vowel, the annotator was prompted to scroll through the extracted video frame-by-frame and identify the point of maximum labial articulation. Points were manually placed at the upper and lower edges of the lip aperture, respectively defined as (i) the boundary between the vermillion border and oral mucosa of the upper lip and (ii) the nearest point on the lower lip. A third point was placed at the oral commissure. The degree of horizontal lip spread was then determined by calculating the horizontal distance between the oral commissure and the plane intersecting with the upper and lower lip aperture points. Although coronal-view video was also recorded, only the sagittal-view video is considered at present.

3.4 Acoustic Results

Figure 3.2 provides normalized vowel formant measurements for all speakers in the Chicago dataset. It is observed that, like the speakers in Chapter 2, speakers in Chicago exhibit fronting of the vowel /u/, such that /u/ occupies a wide range of values for F2 in the upper region of the vowel space, with a mean of 1612 Hz. Unlike the speakers from California and South Carolina, however, /o/ remains backed and exhibits the lowest mean F2 of all the vowels measured for this experiment, at 1267 Hz. For the majority of speakers in this dataset, /æ/ is relatively raised, with a mean F1 of 796 Hz. Most relevant to the present analysis, the speakers in this study exhibit a mean F2 for /ɑ/ of 1517 Hz and a mean F2 for /ɔ/ of 1347 Hz, indicating that both of these vowels are quite fronted for Chicagoans. However, individuals vary in the strength of the contrast between these vowels, as will be discussed below.

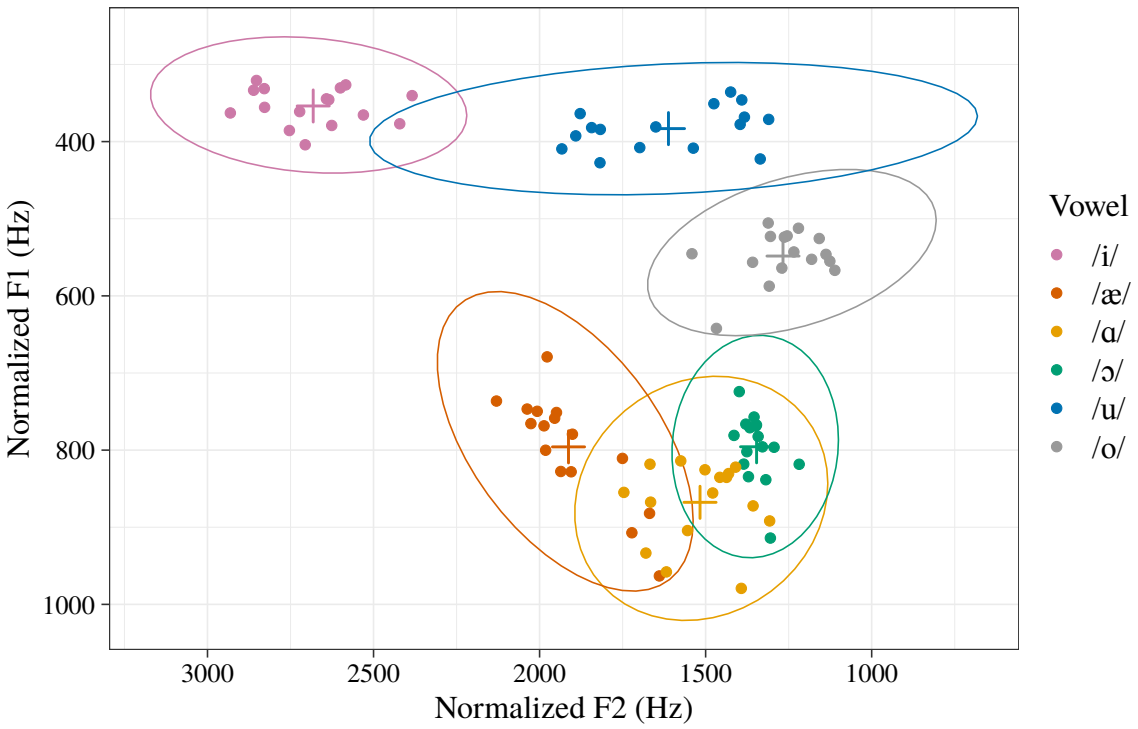


Figure 3.2: Normalized mean formant measurements for all Chicago speakers, normal speech task. Individual points indicate vowel category means for each speaker; cross marks indicate group means. Ellipses indicate 95% confidence intervals.

3.4.1 Participation in the Northern Cities Shift

Given recent findings suggesting that the NCS is in decline among younger speakers, as well as the wide age range of participants in this study, it is useful to determine the extent to which individual participants in this study do or do not participate in the shift. In *The Atlas of North American English*, Labov, Ash, and Boberg (2006) lay out four metrics for establishing the isoglosses that define the geographic extent of the NCS. These criteria, two of which are applied in the present study, are given in (5):

(5) a. AE1: F1 of /æ/ should be less than 700 Hz.

b. O2: F2 of /ɑ/ should be greater than 1450 Hz.

c. EQ: F1 of /ɛ/ is greater than F1 of /æ/ and F2 of /ɛ/ is less than F2 /æ/.

d. ED: F2 of /ɛ/ minus F2 of /ɑ/ should be less than 375 Hz.

The EQ and ED measures cannot be applied to the speakers in this study because no data were collected for the vowel /ɛ/. These two measures provide a holistic benchmark for participation in the NCS, given that they compare measurements from a vowel representing the earliest stages of the NCS (/æ/ or /ɑ/) with measurements for /ɛ/, which was one of the last vowels to undergo change as a result of the NCS. Thus, speakers meeting these two criteria can be considered to be among the most advanced NCS speakers. However, this dataset does allow for evaluation of the AE1 metric, which gauges the degree of raising of /æ/, and the O2 metric, which provides an indication of the frontedness of /ɑ/.

Figure 3.3 presents each speaker's mean F1 for /æ/, with the AE1 criterion represented by a dashed orange line. Strictly speaking, only one speaker, CHI010, meets this criterion. However, it is observed that the distribution of /æ/ is roughly bimodal, with two groups of speakers producing more and less raised variants of /æ/. The first group, those with the most raised productions of /æ/, exhibit a mean F1 for /æ/ of 769 Hz. While greater than 700 Hz, and therefore not meeting the AE1 criterion, these speakers nevertheless produce /æ/ with an

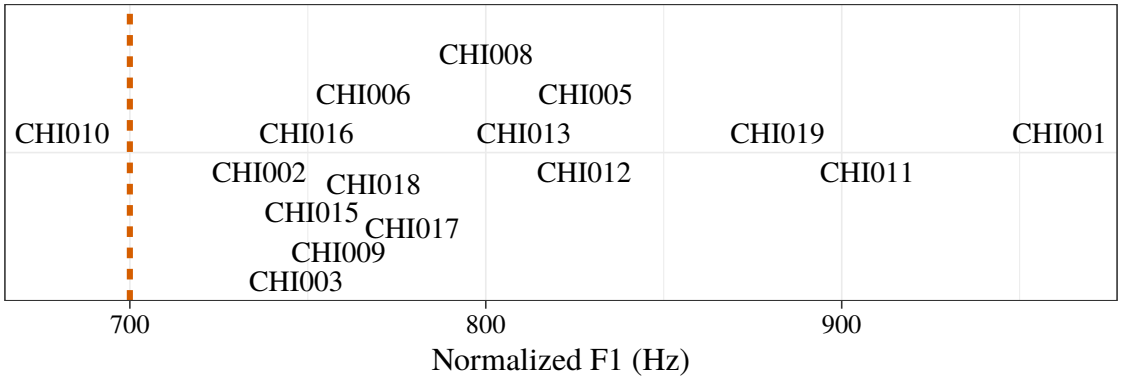


Figure 3.3: Mean F1 of /æ/ for all Chicago speakers, normal speech task. Orange line indicates AE1 criterion; speakers to the left of this line are considered to have highly raised productions of /æ/.

F1 that is substantially lower than what would be expected for speakers outside of the Inland North. For instance, Hillenbrand et al. (1994) find that women from the Midwest produce /æ/ with a mean F1 of 669 Hz, while Hagiwara (1997) finds that women from California produce /æ/ with a mean F1 of approximately 1000 Hz. Only a small number of speakers in this study produce /æ/ with a mean F1 of greater than 800 Hz, and it is plausible that the relatively high F1 observed here is due to the relative formality of speech collected in the lab.

The second group, comprising four speakers, produces /æ/ with a mean F1 of 917 Hz, which is more typical of dialects in which /æ/ remains in the lowest region of the vowel space. As can be observed in Figure 3.2, three of these four speakers also produce /æ/ with a rather low F2 that encroaches on the distribution of other speakers' productions of /ɑ/. This pattern is consistent with studies on the reversal of the NCS, which have shown that younger, college-educated speakers who eschew the NCS adopt a continuous or nasal /æ/

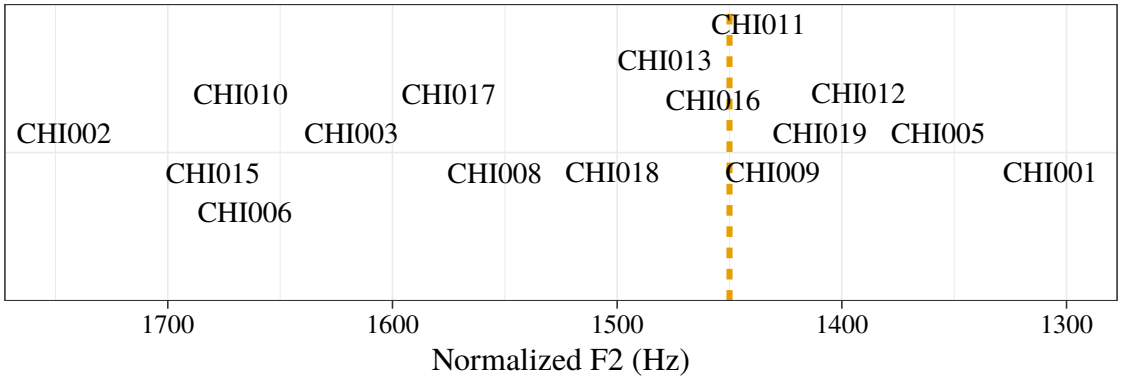


Figure 3.4: Mean F2 of /ɑ/ for all Chicago speakers, normal speech task. Yellow line indicates O2 criterion; speakers to the left of this line are considered to have highly fronted productions of /ɑ/.

system (Wagner et al. 2016), in which /æ/ is raised only in pre-nasal environments (Labov, Ash, and Boberg 2006).[5] In addition, the retraction of /æ/ is a feature of the "third dialect" of North American English (Labov 1991, 1994), which encompasses regions not exhibiting the Northern Cities Shift or the Southern Shift; Wagner et al. (2016) argue that younger speakers in Lansing, Michigan have begun to orient toward a vowel system typical of the third dialect. Indeed, the speakers in this study with the least raised and most retracted productions of /æ/ are among the youngest participants in the study, with a mean age of 21.7 years (vs. 52.2 years for raised-/æ/ speakers).

In contrast, however, the majority of participants in this study do achieve the O2 criterion. The mean F2 of /ɑ/ for each participant, along with the O2 criterion, is presented in Figure 3.4. Ten participants exhibit a mean F2 of greater than 1450 Hz, while six fall below

5. The only wordlist item in which /æ/ appears in a pre-nasal context is *dan*; all other items contain oral codas. Thus, any raising that might occur for pre-nasal /æ/ would not be reflected in the mean F1 values reported here.

this criterion. Even though not all participants produce /ɑ/ with an F2 greater than 1450 Hz, however, none of the participants produces /ɑ/ with a mean F2 of less than 1300 Hz, indicating that /ɑ/ is in general quite fronted for these speakers. CHI002, the speaker with the most fronted production of /ɑ/, exhibits an extremely high mean F2 of nearly 1800 Hz.

Based on these metrics, it appears that the participants in this study vary in the extent to which they exhibit the NCS. While the majority of speakers exhibit at least some degree of /æ/-raising, there are several younger speakers in this sample who produce pre-oral /æ/ with a low F1, which is more typical of nasal or continuous /æ/ systems found elsewhere in North America. While all speakers produce /ɑ/ with a fairly high F2, there is a range of variability, with some speakers producing /ɑ/ with an F2 of 1300 Hz, and others producing /ɑ/ with an extremely high F2 close to 1800 Hz.

3.4.2 Maintenance of the cot-caught Contrast

Beyond /æ/-raising and /ɑ/-fronting, another hallmark feature of the NCS is the fronting of /ɔ/. Because the fronting of /ɔ/ follows the fronting of /ɑ/, speakers typically retain the contrast between these two vowels. As noted above, however, several articulatory strategies exist when it comes to fronting a back round vowel such as /ɔ/. Havenhill and Do (2018) find that speakers from Michigan vary in the degree to which they maintain the contrast between /ɑ/ and /ɔ/, as a result of adopting differing articulatory configurations. They find that speakers who produce a smaller acoustic distinction between /ɑ/ and /ɔ/ do so by abandoning one of the articulatory gestures that typically distinguishes /ɔ/ from /ɑ/, i.e., lip rounding and tongue backing.

As was found for speakers from Metro Detroit, speakers from Chicago vary in the degree to which /ɑ/ and /ɔ/ differ acoustically. Figure 3.5 presents formant measurements for /ɑ/ and /ɔ/ as a kernel density estimation plot, which shows the distribution of tokens containing /ɑ/ and /ɔ/ in the F1 × F2 space. A wide range of distributions for /ɑ/ and /ɔ/ is observed. For

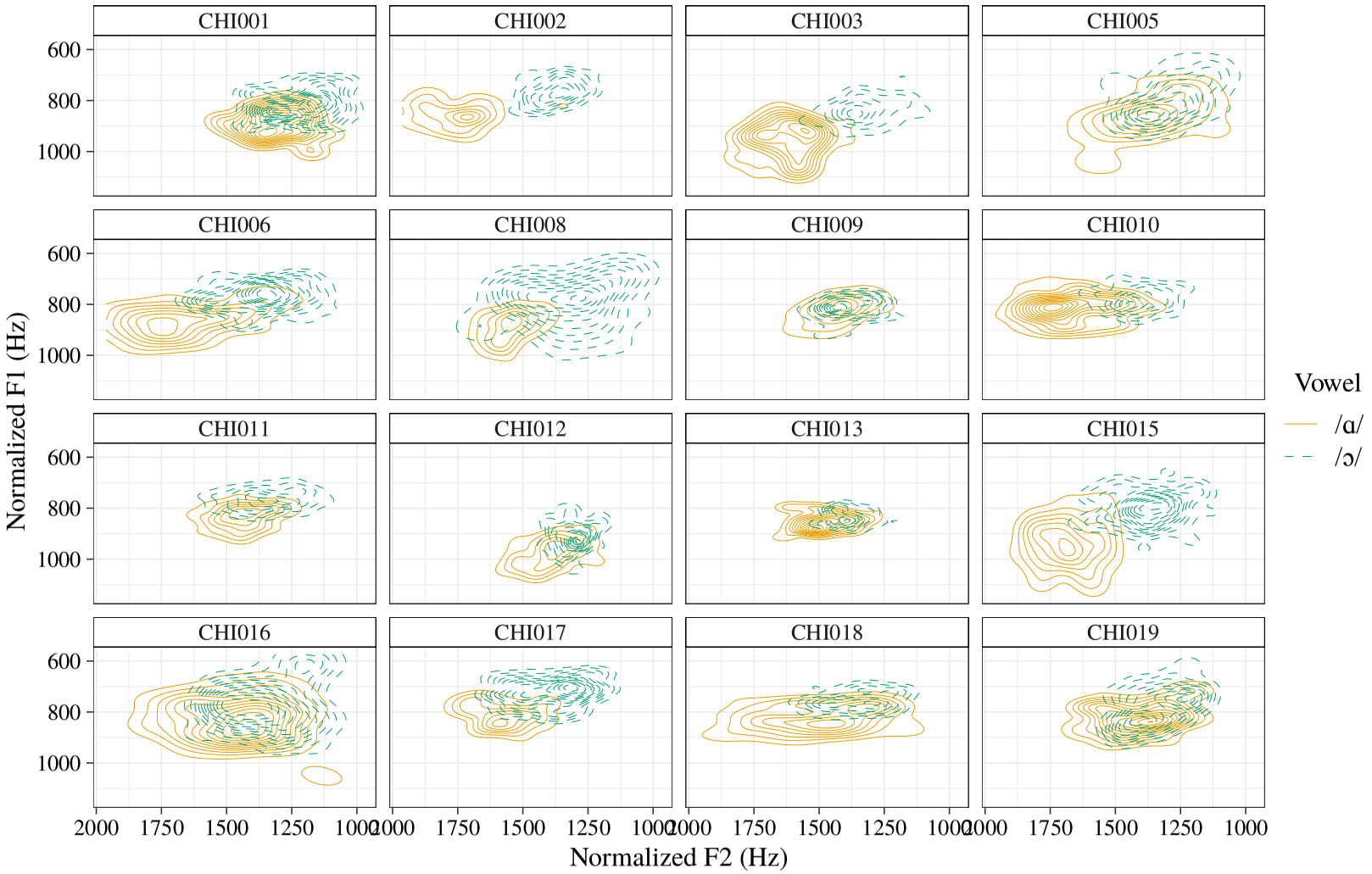


Figure 3.5: Kernel density estimation plot for /ɑ/ and /ɔ/ in normal speech task, all participants.

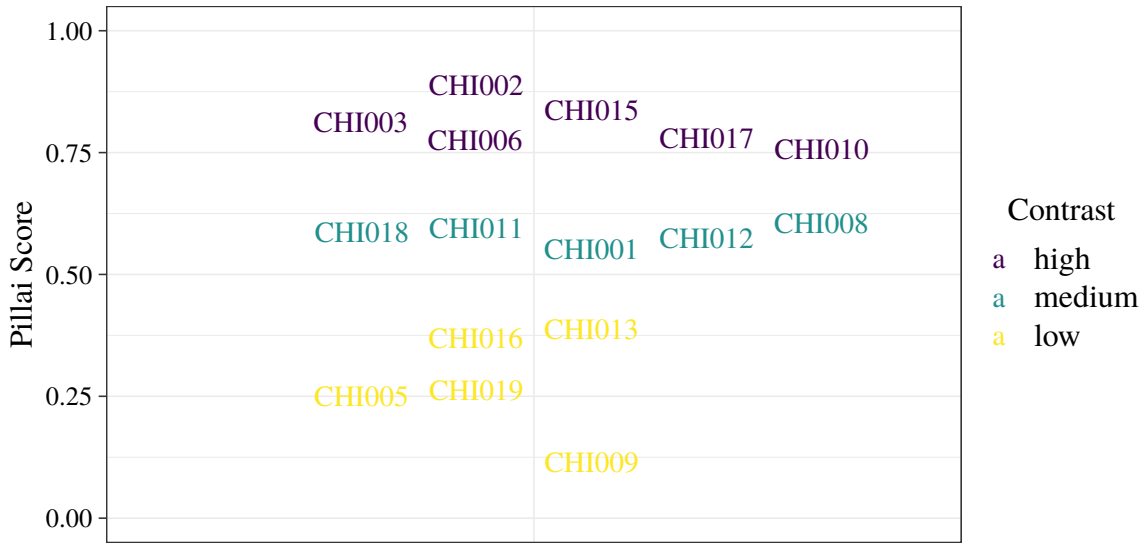


Figure 3.6: Pillai scores for normal speech task, all participants. Speakers with a higher score exhibit a greater degree of acoustic contrast between /ɑ/ and /ɔ/.

instance, CHI002 produces these vowels with little or no overlap, while CHI009 exhibits nearly complete overlap, suggesting that he has merged the two categories. For speakers including CHI013 and CHI018 (and to some extent CHI010) the distribution for /ɔ/ is a subset of that for /ɑ/, with /ɑ/ occupying a relatively wide range of values along the F2 dimension.

In order to quantify this variability, a Pillai score was calculated for each speaker. A Pillai score is the output of a multivariate ANOVA, which takes into account both the distance between the vowel means and the overlap between the vowel distributions. This measure was first used for sociophonetic work by Hay, Nolan, and Drager (2006), and has since found a wide range of applications in the literature. The output of this model is a numeric value between 0 and 1, with a score of 0 indicating completely identical distributions and

a score of 1 indicating completely distinct distributions. In addition, this model incorporates the preceding and following segments as predictors, in order to filter out the effects of phonological environment.

The Pillai score for each speaker is presented in Figure 3.6. The participant with the highest Pillai score, indicating the greatest difference between /ɑ/ and /ɔ/, is CHI002, with a score of 0.889. This is unsurprising, given that this participant also exhibits the greatest degree of /ɑ/ fronting, and is among the speakers with the most raised productions of /æ/. The participant with the lowest Pillai score of 0.116 is CHI009, who was observed in Figure 3.5 to have nearly complete overlap of the vowel categories. Other participants fall elsewhere within this range, and can be roughly divided into three distinct groups. Six speakers receive a score of greater than 0.75, indicating a strong contrast between the two vowels. Five participants receive a score between 0.5 and 0.65 (a moderate degree of contrast) and five speakers exhibit a weak /ɑ/-/ɔ/ contrast with scores well below 0.5.

As noted above, recent studies on the reversal of the NCS have shown that younger, college-educated speakers exhibit a reorganization of the vowel space typically observed for speakers in the Inland North, with /æ/ undergoing an allophonic split and /ɑ/ reversing its typically fronted realization. Moreover, Savage et al. (2016) has found that fronted productions of /ɑ/ have become socially stigmatized in the minds of younger listeners, with participants describing fronted /ɑ/ as "ignorant and annoying." If /ɑ/ begins to back as a result of this stigma, it is reasonable to predict that, unless /ɔ/ also undergoes backing, the contrast between /ɑ/ and /ɔ/ will be reduced. Figure 3.7 presents the Pillai score for each speaker plotted against their age. This plot shows that, in general, older speakers exhibit a higher degree of contrast between /ɑ/ and /ɔ/ than younger speakers. All of the speakers with a Pillai score of greater than 0.75 are over 50 years old, and the majority of speakers with a Pillai score below 0.75 are under the age of 40. There are three older speakers in this sample who are exceptions to this trend: speakers CHI016 and CHI005, who exhibit

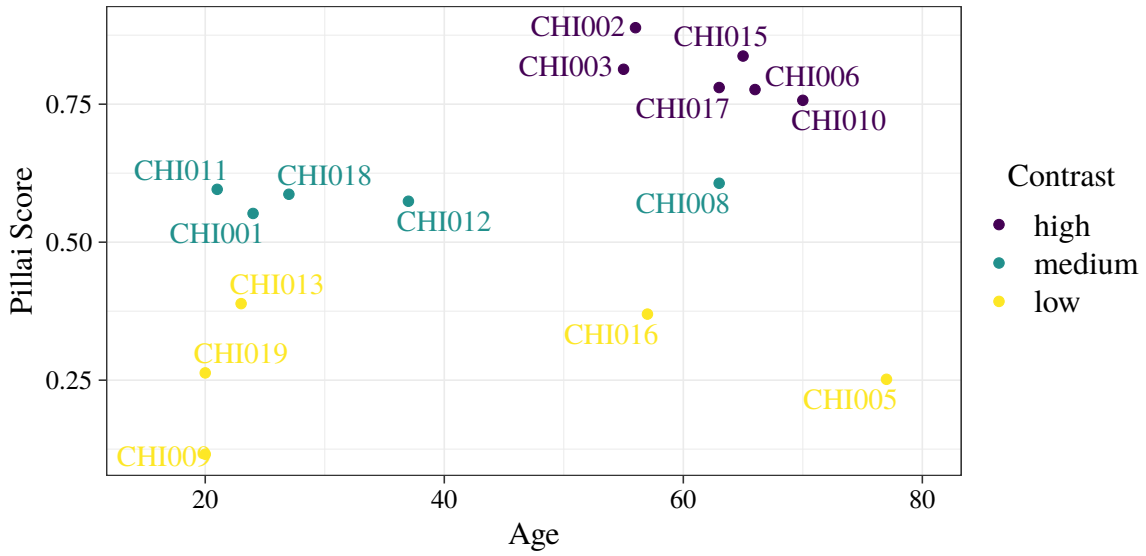


Figure 3.7: Pillai scores for normal speech task by age.

very low Pillai scores below 0.375, and speaker CHI008, who exhibits a moderate degree of contrast between /ɑ/ and /ɔ/ with a Pillai score of 0.607. CHI005 and CHI016 are the two speakers in this study who have spent the most time living outside the Chicago area, which may go some way toward explaining why their COT-CAUGHT contrasts (or lack thereof) do not fit the expected pattern.

3.4.3 Enhancement of the cot-caught Contrast

As described in Section 3.3.3, participants were also asked to produce words containing the target vowels, /ɑ/ and /ɔ/, in a task designed to elicit careful speech. Figure 3.8 shows each participant's Pillai score in both the normal and careful speech tasks.[6] For the majority of

6. CHI009 did not complete the careful speech task due to time constraints, which is unfortunate because those data would be informative with respect to the status of the COT-CAUGHT merger for this speaker.

the participants in this study, production of /ɑ/ and /ɔ/ in the careful speech task results in an increase in the separation of these vowels in the F1 × F2 space, as indicated by an increase in Pillai score. For speakers who were in the high contrast group in the normal speech task, the increase in Pillai score in the careful speech task is relatively small. This is not especially surprising given that /ɑ/ and /ɔ/ are already quite distinct for these speakers. More interestingly, nearly all of the speakers in the medium and low contrast groups drastically increase the degree of contrast between these two vowels in the careful speech task, providing strong evidence that /ɑ/ and /ɔ/ are phonologically distinct for these speakers. For instance, CHI013 exhibits a relatively low score of 0.389 in the normal speech task, but a high score of 0.707 in the careful speech task.

Two participants, CHI011 and CHI003, however, produce these vowels in a *less* distinct manner in the careful speech task, indicated by dashed lines in Figure 3.8. The distribution of /ɑ/ and /ɔ/ in normal and careful speech for these two speakers is given in Figure 3.9. For CHI003, both /ɑ/ and /ɔ/ exhibit a significantly higher F1 in the careful speech task than in the normal speech task. This result can be attributed to this speaker speaking with more extreme jaw movement in the careful speech task. This more extreme articulation results in low vowels such as /ɑ/ and /ɔ/ being produced with a lower jaw, which has the effect of raising F1. Thus, while this speaker does appear to adopt more effortful articulations in the careful speech task, there is no increase in the spectral distance between the two vowels. For CHI011, the decrease in Pillai score is the result of convergence between the vowels /ɑ/ and /ɔ/. The mean of /ɑ/ is both higher and backer in the careful speech task, while the range of /ɔ/ has decreased, such that no tokens appear in the highest and backest region of this vowel's distribution.

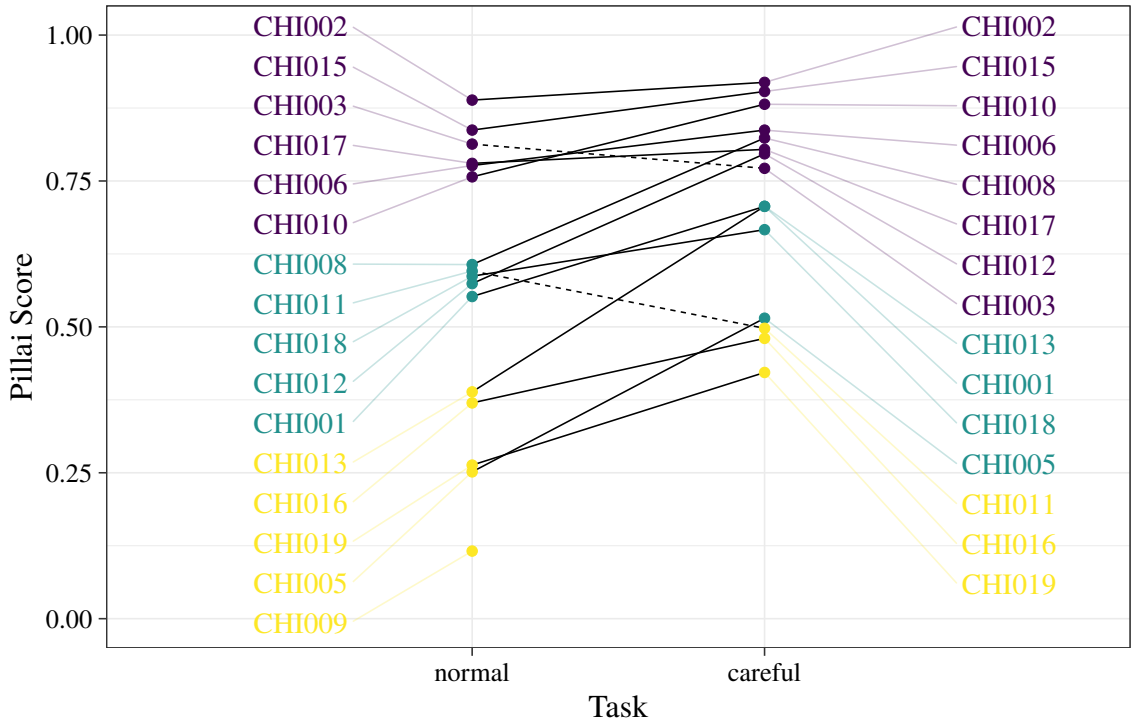


Figure 3.8: Pillai score by task, all participants. Solid line indicates increase in Pillai score for careful speech task, dashed line indicates decrease.

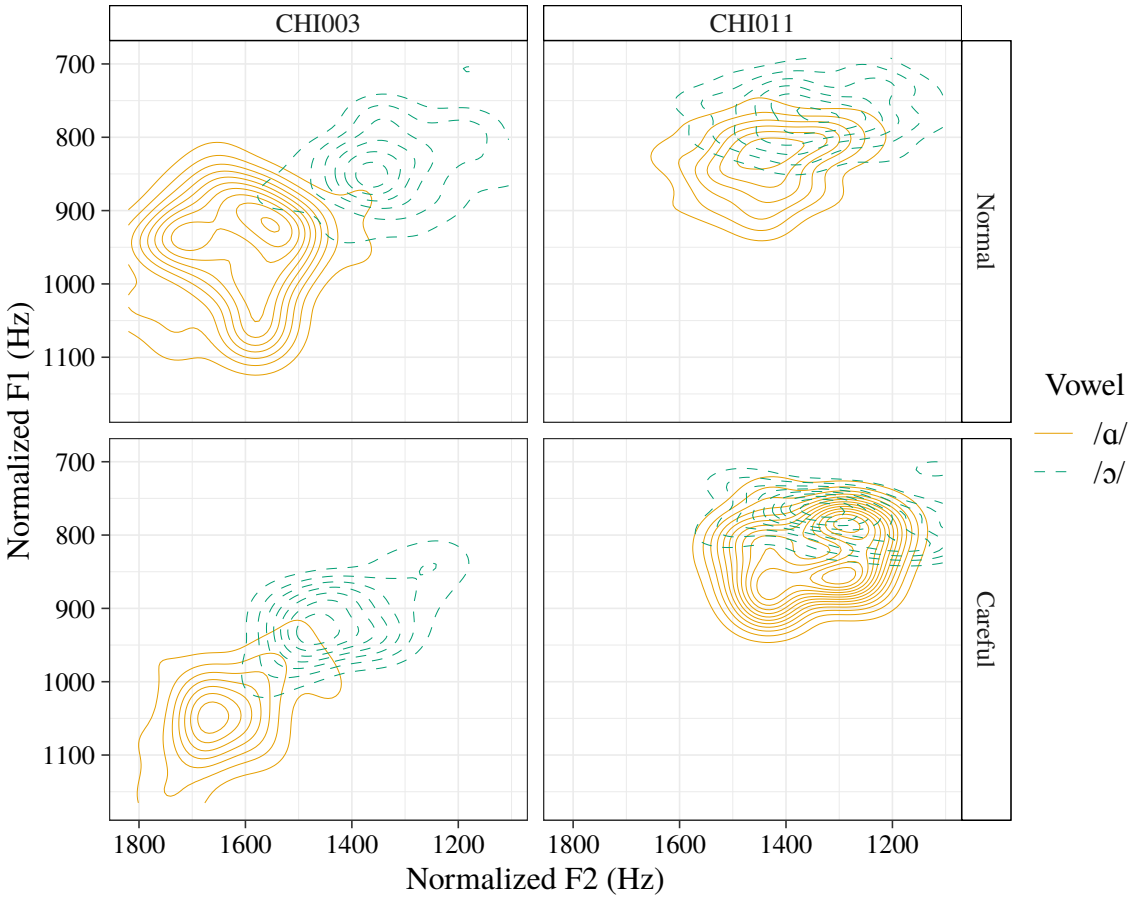


Figure 3.9: Distribution of /ɑ/ and /ɔ/ for CHI003 and CHI011 in normal and careful speech.

3.5 Articulatory Results

Section 3.4 demonstrated that Chicagoans vary in the extent to which they maintain the contrast between /ɑ/ and /ɔ/. While some speakers exhibit a strong acoustic contrast between these vowels, such that the distributions for /ɑ/ and /ɔ/ exhibit little or no overlap, other speakers produce these vowels with almost complete overlap. Given that fronting of the vowel /ɔ/ can be achieved by multiple articulatory strategies, it is possible that these differences in acoustic contrast are the result of differences in articulatory configuration, as was found for Michiganders by Havenhill and Do (2018). In this section, articulatory data from ultrasound tongue imaging and lip video is presented. It is shown that speakers with the highest degree of acoustic contrast between /ɑ/ and /ɔ/ produce these vowels with differences in both tongue position and lip rounding, while speakers with a smaller degree of acoustic contrast tend to distinguish the vowels with differences in lip rounding alone. Only one speaker distinguishes /ɔ/ from /ɑ/ solely in terms of tongue position, lending support to the prediction that this pattern of articulation will be rare or dispreferred. Further, it is shown that in careful speech, the majority of speakers increase the magnitude of the lip rounding difference between /ɑ/ and /ɔ/.

3.5.1 Tongue Position

Analysis of the ultrasound tongue imaging data was conducted with polar smoothing spline ANOVA (SS ANOVA), as described in Section 3.3.4. In this case, SS ANOVA is used to determine whether the tongue shape and position differ significantly for the vowels /ɑ/ and /ɔ/. Smoothing spline estimates for CHI010 are presented in Figure 3.10. This plot shows smoothing splines for each of the vowels measured in this study, along with 99% Bayesian confidence intervals. As expected, the vowel with the highest point of constriction and most

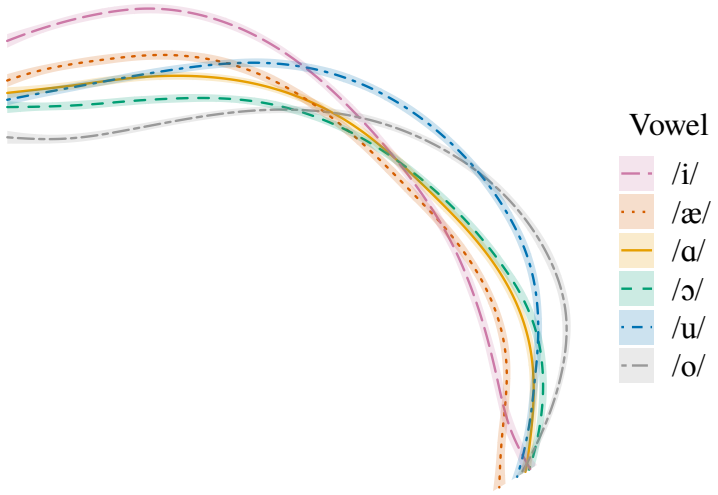


Figure 3.10: Smoothing spline estimates for CHI010, all vowels. Tongue front is to the left, tongue root is to the right. Shading indicates 99% confidence interval. Splines rotated 16.2° clockwise to orient the occlusal plane horizontally.

advanced tongue root is /i/, while the vowel with the greatest degree of pharyngeal constriction is /o/. Most relevant to the present analysis, this speaker produces /ɑ/ and /ɔ/ with distinct tongue shapes and positions, and the lack of overlap between the shaded confidence intervals for these vowels shows that this difference is statistically significant.

SS ANOVA tongue contours for the vowels /ɑ/ and /ɔ/ for all sixteen speakers in the dataset are presented in Figure 3.11. As with the smoothing spline estimates presented in Figure 3.10, the shaded regions surrounding the smoothing splines indicate 99% Bayesian confidence intervals; where the intervals for /ɑ/ and /ɔ/ overlap, the difference between the tongue contours for /ɑ/ and /ɔ/ is not statistically significant. This plot reveals that eight speakers exhibit a significant difference between /ɑ/ and /ɔ/ in tongue position, while eight participants produce the two vowels with identical tongue positions. Thus, the ultrasound

data suggest that there is a wide range of variability with respect to whether or not /ɑ/ and /ɔ/ are contrasted in terms of tongue position.

3.5.2 Lip Rounding

In contrast, there is far less variability when it comes to lip rounding measurements. Figure 3.12 presents normalized lip spread measurements for /ɑ/ and /ɔ/ for the same sixteen speakers. It will be noted that while two speakers (CHI009 and CHI013) do not produce /ɑ/ and /ɔ/ with a signficiant difference in lip rounding, the vast majority (fourteen of sixteen) retain the rounding contrast between these two vowels.

3.5.3 Summary of Articulatory Results

The articulatory and acoustic results for the normal speech task are summarized in Table 3.2, with speakers categorized according to their articulatory strategy. This table shows that half of the participants in this study do distinguish between /ɑ/ and /ɔ/ in terms of both lip rounding and tongue position. For most such speakers, the acoustic contrast between /ɑ/ and /ɔ/ is relatively high, such that these speakers exhibit a mean Pillai score of 0.752, the highest among all groups. The next largest group is those who distinguish between /ɑ/ and /ɔ/ with a difference in lip rounding, but without a difference in tongue position. Two speakers appear to exhibit a merger of the two vowels,[7] Finally, only a single speaker maintains the /ɑ/ and /ɔ/ contrast through a difference in tongue position alone.

7. One may wonder why CHI019 has been categorized as exhibiting a merger of /ɑ/ and /ɔ/, when this participant does in fact exhibit a significant difference in lip spread. This determination was made based on three factors: a) this participant's exceptionally low Pillai score, b) her behavior in the careful speech task, where this difference in lip rounding disappears, and c) her performance in the perception task presented in Chapter 4, where she does not correctly identify these vowels above chance. This speaker's behavior is therefore a fairly typical case of near-merger; Labov, Karen, and Miller (1991) observe that speakers with a near-merger of two phonemes generally produce the two sounds with a relatively small phonetic difference, that this phonetic difference is reduced in more careful speech styles, and that the two sounds are judged in perception tasks to be the same.

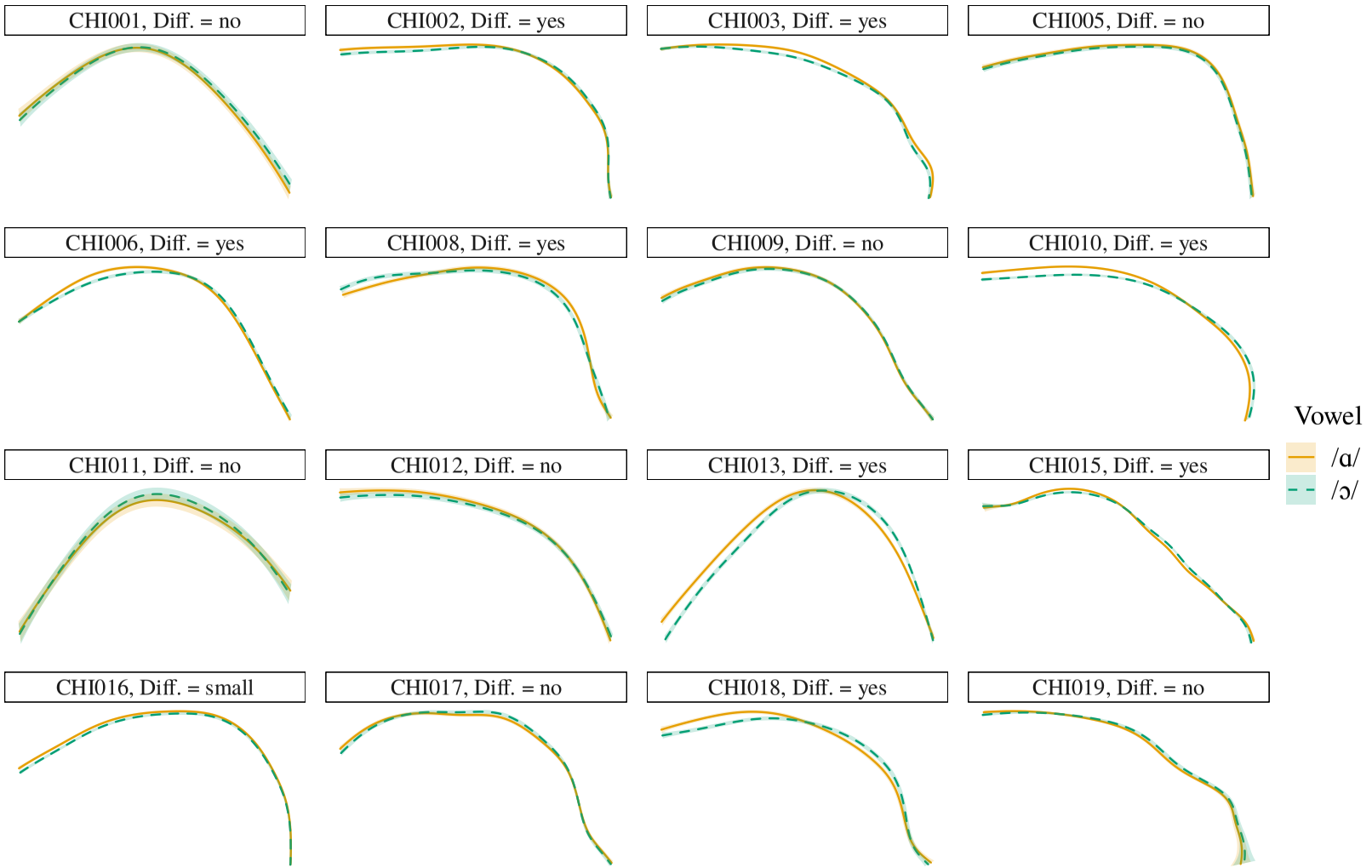


Figure 3.11: Smoothing spline estimates for /ɑ/ and /ɔ/ for Chicago speakers, normal speech task. Shading indicates 99% Bayesian confidence interval. *Diff.* indicates whether there is a significant difference (non-overlapping portion) between the two confidence intervals.

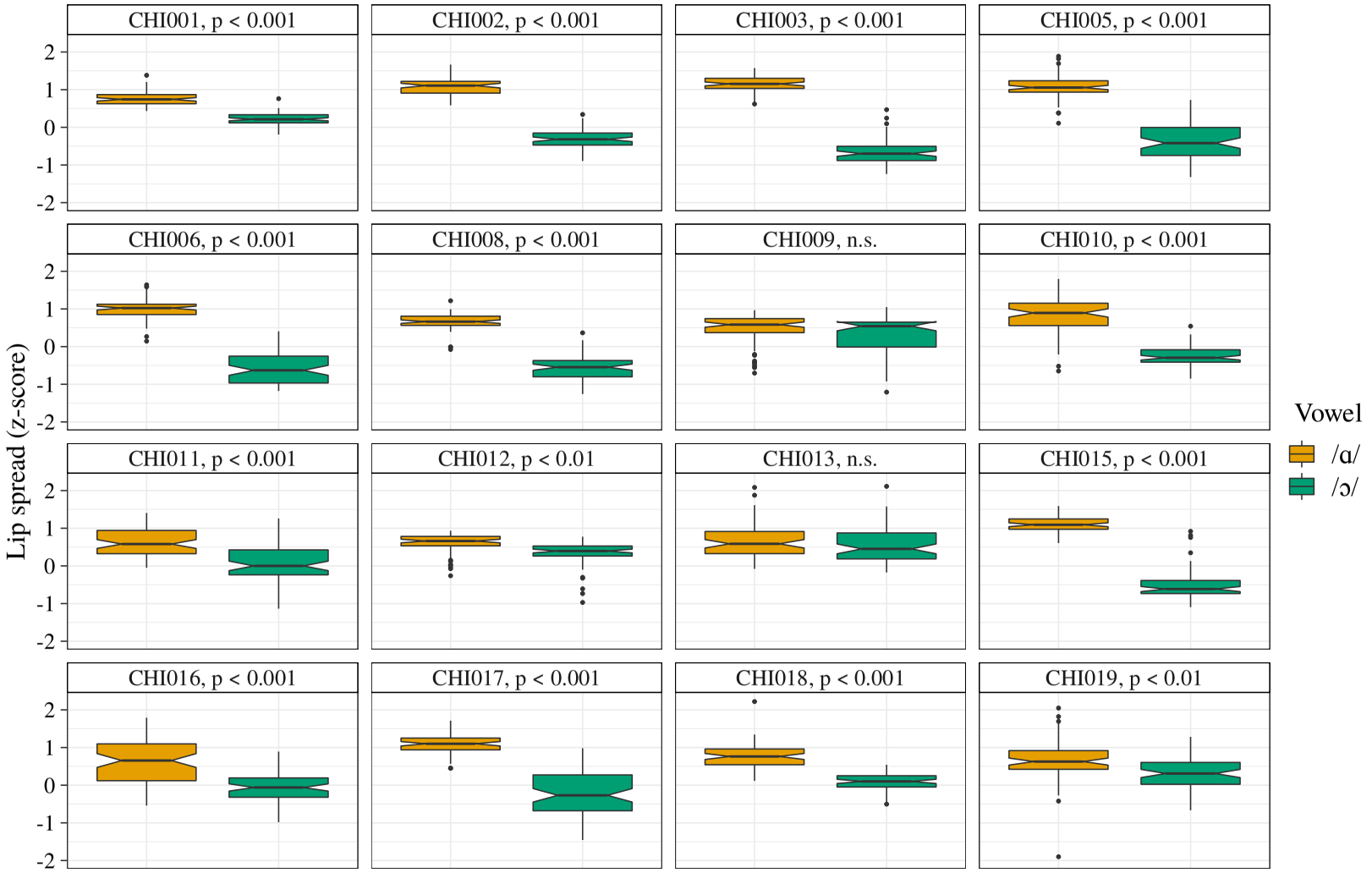


Figure 3.12: Lip spread measurements for Chicago speakers. Smaller values indicate increased rounding.

Table 3.2: Summary of acoustic and articulatory results for Chicago speakers, normal speech task. For tongue difference, *yes* indicates that the confidence intervals for /ɑ/ and /ɔ/ do not overlap for some portion of the tongue. For lip spread difference, *yes* indicates that /ɑ/ and /ɔ/ differ significantly ($p < 0.05$).

Speaker ID	Pillai score	Tongue difference	Lip spread difference	Strategy
CHI002	0.889	yes	yes	Both
CHI015	0.837	yes	yes	Both
CHI003	0.813	yes	yes	Both
CHI006	0.776	yes	yes	Both
CHI010	0.757	yes	yes	Both
CHI008	0.607	yes	yes	Both
CHI018	0.587	yes	yes	Both
CHI017	0.780	no	yes	Lip
CHI011	0.596	no	yes	Lip
CHI012	0.574	no	yes	Lip
CHI001	0.552	no	yes	Lip
CHI016	0.370	small	yes	Lip
CHI005	0.252	no	yes	Lip
CHI013	0.389	yes	no	Tongue
CHI019	0.263	no	yes	Merged
CHI009	0.116	no	no	Merged

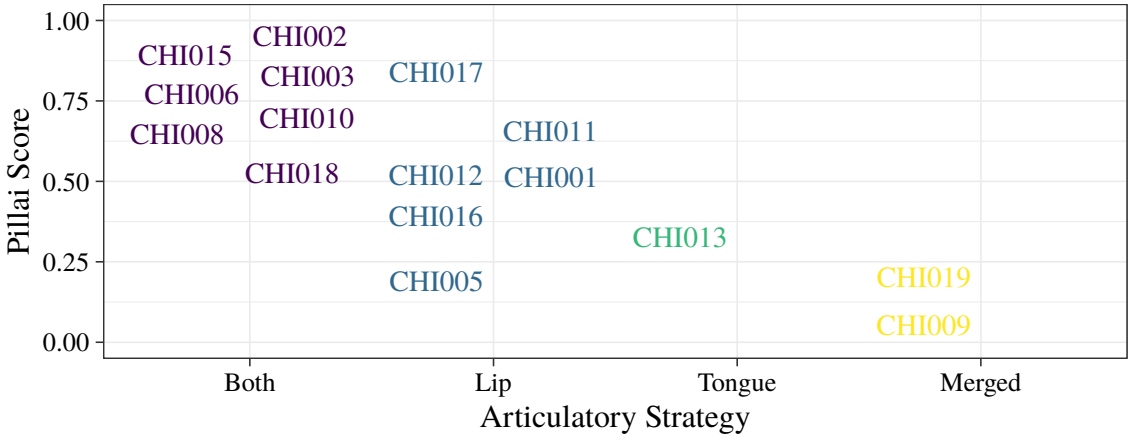


Figure 3.13: Pillai scores by articulatory strategy, normal speech task.

Finally, Figure 3.13 presents the Pillai score for each speaker, with speakers grouped by articulatory pattern. It is observed that, as was found by Havenhill and Do (2018) for Michiganders, the speakers with the highest Pillai score are those who produce an articulatory contrast between /ɑ/ and /ɔ/ with both tongue position and lip rounding. For these speakers, the mean Pillai score is 0.752. For speakers who maintain a difference between /ɑ/ and /ɔ/ in terms of lip rounding alone, the mean Pillai score is somewhat lower, at 0.521. The speaker who produces a contrast between /ɑ/ and /ɔ/ with a difference in tongue position exhibits a Pillai score of 0.389. Finally, the two speakers in this study who do not produce a distinction between /ɑ/ and /ɔ/ exhibit a mean Pillai score of 0.19 A one-way ANOVA indicates that articulatory strategy is a signficant predictor of Pillai score ($F_{4,95} = 8.71$, $p < .01$), but a Tukey post hoc test reveals that not all groups differ significantly from one another. Notably, the difference between the "Both" and "Lip" groups is not significant, nor is the difference between the "Lip" and "Tongue" groups.

3.5.4 Articulatory Strategies for cot-caught Contrast Enhancement

As observed in Figure 3.8, the majority of speakers in this study increased the spectral contrast between /ɑ/ and /ɔ/ when speaking in a careful manner. This section considers the articulatory strategies used to enhance this contrast, testing the hypothesis that speakers optimize their speech not only for auditory perceptibility, but also for visual perceptibility. As discussed in Chapter 1, teleological models of phonology predict that under adverse communicative conditions, speakers will actively enhance their speech production patterns based on the listener's perceptual requirements. Given that listeners are sensitive not only to auditory perceptual cues but also visual speech cues, it is predicted that in careful speech, speakers will increase the degree to which /ɑ/ and /ɔ/ differ visually through an increase in the magnitude of the labial gestures involved for these vowels. That is, speakers are predicted to increase the degree of lip spread for /ɑ/ and to decrease the degree of lip spread (i.e., increase lip protrusion) for /ɔ/.

Figure 3.14 presents SS ANOVA tongue contours for /ɑ/ and /ɔ/, as produced in the careful speech task. A number of differences are observed when comparing these tongue contours to those given in Figure 3.11. Several speakers exhibit a tongue position distinction in the careful speech task who did not do so in the normal speech task, including all but two of the speakers who were categorized as distinguishing between these vowels in terms of lip rounding alone. CHI019, who exhibits a near-merger of /ɑ/ and /ɔ/, also produces /ɔ/ with an increased pharyngeal constriction relative to /ɑ/. Among speakers who distinguished /ɔ/ from /ɑ/ in terms of both tongue position and lip rounding in the careful speech task, the majority continue to produce a tongue position contrast in the careful speech task, as would be expected. Curiously, however, two speakers who distinguished these vowels through tongue position in the normal speech task do not do so in the careful speech task. These speakers include CHI002, who was the speaker with the highest Pillai score in both

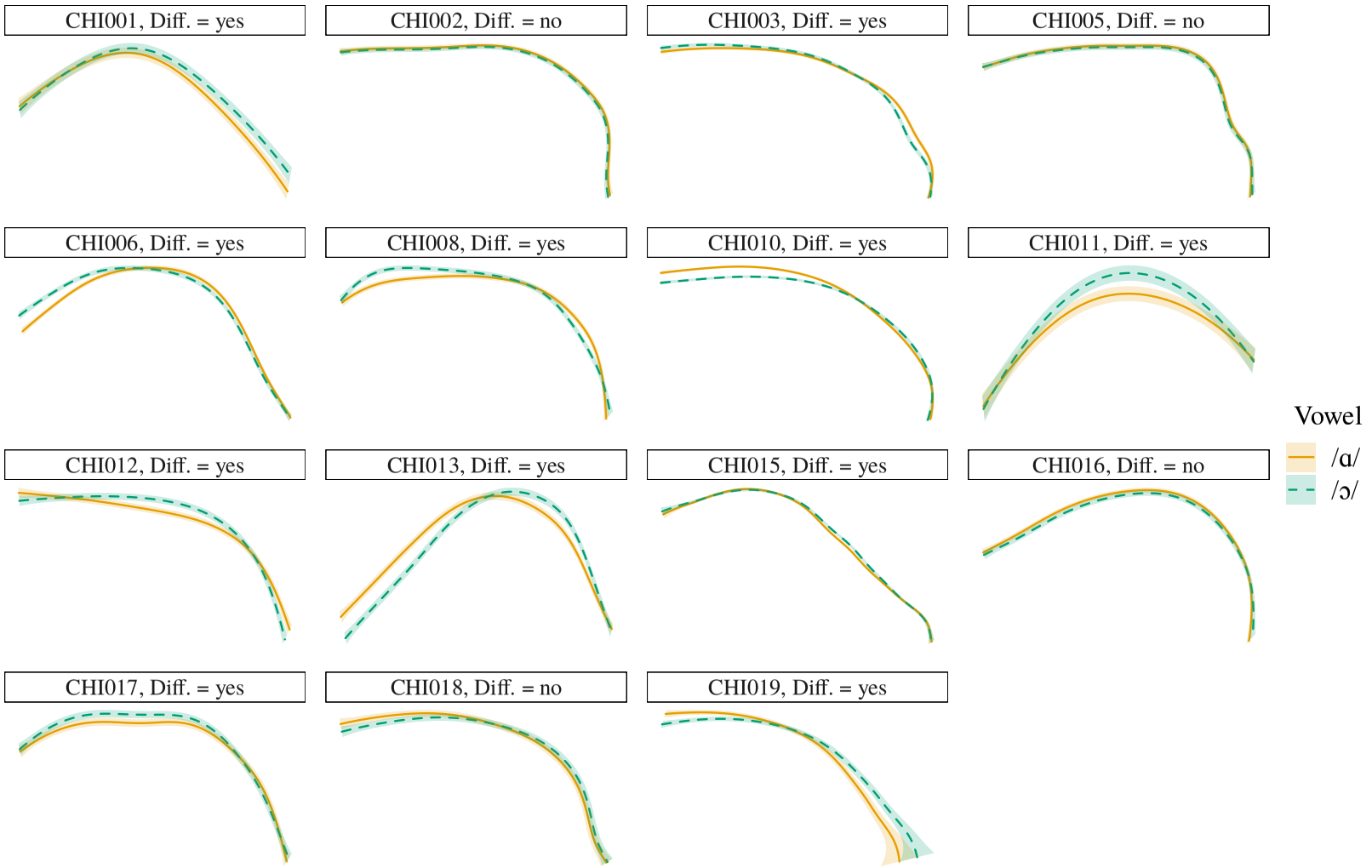


Figure 3.14: Smoothing spline estimates for /ɑ/ and /ɔ/, careful speech task. Shading indicates 99% Bayesian confidence interval. *Diff.* indicates whether there is a significant difference between the two tongue splines, i.e., a non-overlapping segment of the confidence intervals.

tasks, and CHI018, who was the speaker with the lowest Pillai score among speakers producing both tongue and lip distinctions between /ɑ/ and /ɔ/. While both speakers produce these vowels with a greater degree of acoustic separation in the careful speech task, the increase in Pillai score is relatively small: 0.030 for CHI002 and 0.080 for CHI018.

Figure 3.15 presents lip spread measurements for Chicago speakers in both the normal and careful speech tasks. Several patterns emerge. First, some speakers are observed to decrease the degree of lip spread for /ɔ/, which is the result of increasing the degree of lip rounding. Notably, CHI013, who was the only non-merged speaker in the normal speech task not to produce a lip rounding contrast between /ɑ/ and /ɔ/, significantly increases the degree of lip rounding for /ɔ/. Other speakers who exhibit such an increase include CHI002, who was shown in Figure 3.14 not to produce a tongue position contrast between /ɑ/ and /ɔ/ in careful speech. Given that the Pillai score for this speaker still showed an increase in the careful speech task, it appears that this increase in lip rounding overcomes the lack of distinction in tongue position. On the other hand, some speakers do not significantly increase the degree of lip rounding for /ɔ/, but do increase the degree of lip spread for /ɑ/. Because such an increase will also serve to enhance the visual distinctiveness of these two vowels, speakers who increase the degree of lip spread for /ɑ/ can also be taken as support for the hypothesis that speakers optimize their speech (in part) for visual perceptibility. Speakers who significantly increase the degree of lip spread for /ɑ/ include CHI008 and CHI016.

Table 3.3 presents a summary of the articulatory results for the careful speech task. Each speaker is listed along with their articulatory classification from the normal speech task, the change in their Pillai score, whether they produced a significant tongue position distinction, and the direction and significance of any change in lip spread for /ɑ/ and /ɔ/. The tongue contrast column shows that in the careful speech task, all but four speakers produced a tongue position distinction between the two vowels. Four speakers who did not produce a tongue position distinction in normal speech did so in careful speech, as did CHI019, who

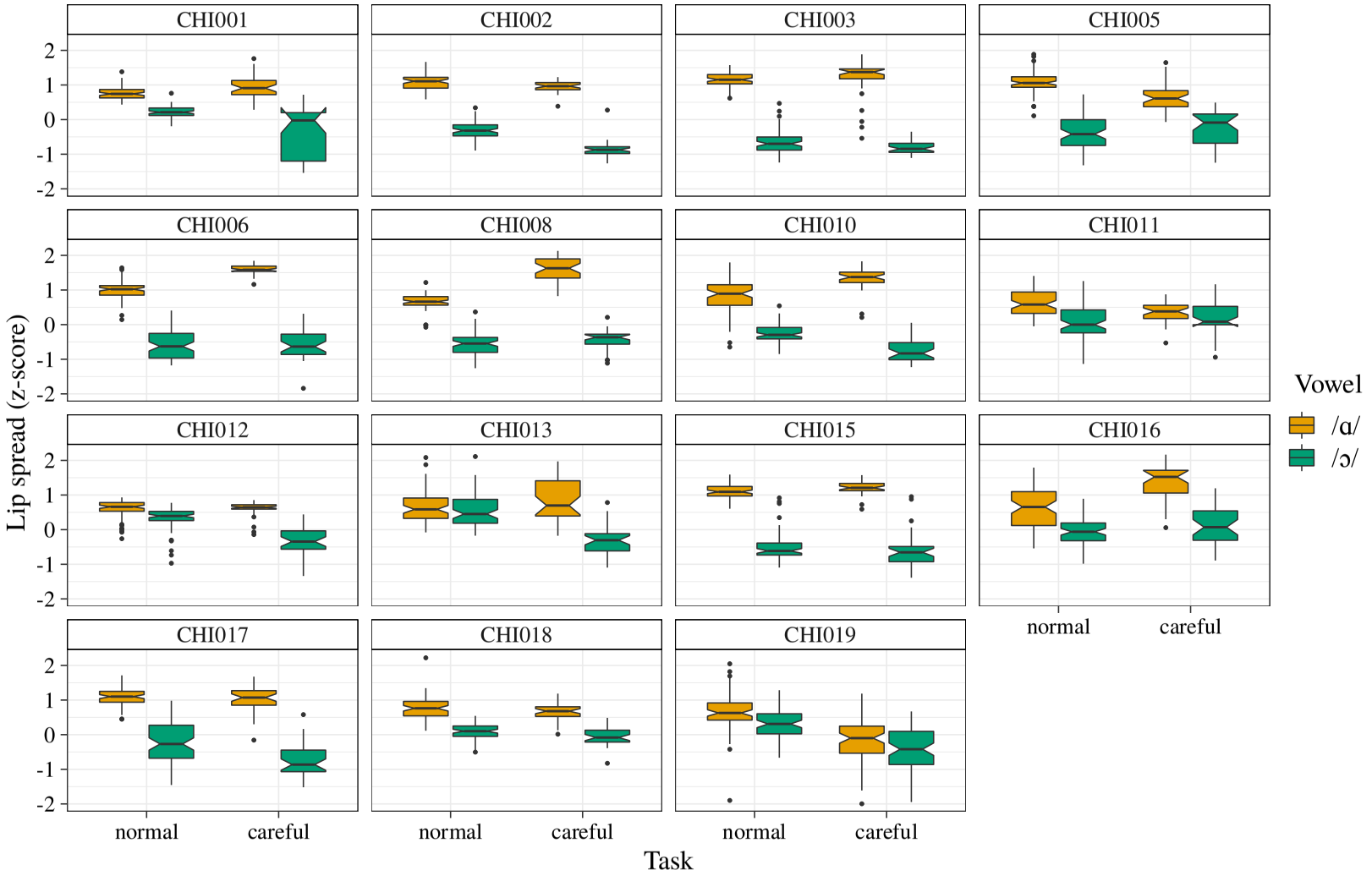


Figure 3.15: Lip spread measurements for Chicago speakers, normal and careful speech tasks. Smaller values indicate increased rounding.

Table 3.3: Summary of contrast enhancement strategies for Chicago speakers. Boldface cells indicate an articulatory enhancement not observed in normal speech.

Speaker ID	Articulatory Strategy	Change in Pillai score	Tongue contrast	Change in /ɑ/ spread	Change in /ɔ/ spread
CHI002	Both	0.030	no	Decrease	Decrease
CHI015	Both	0.066	yes	**Increase**	no
CHI003	Both	-0.042	yes	no	**Decrease**
CHI006	Both	0.061	yes	**Increase**	no
CHI010	Both	0.124	yes	**Increase**	**Decrease**
CHI008	Both	0.217	yes	Increase	Increase
CHI018	Both	0.080	no	no	**Decrease**
CHI017	Lip	0.024	**yes**	no	**Decrease**
CHI011	Lip	-0.098	**yes**	Decrease	no
CHI012	Lip	0.222	**yes**	no	**Decrease**
CHI001	Lip	0.154	**yes**	**Increase**	**Decrease**
CHI016	Lip	0.110	no	**Increase**	no
CHI005	Lip	0.263	no	Decrease	no
CHI013	Tongue	0.318	yes	no	**Decrease**
CHI019	Merged	0.159	**yes**	Decrease	Decrease

exhibits a near-merger of /ɑ/ and /ɔ/. CHI005 and CHI016 did not produce a tongue position distinction in either task. While CHI002 and CHI018 produced a tongue position distinction in the normal speech task, they did not do so in the careful speech task. This loss of tongue contrast is accompanied by a comparatively small increase in acoustic contrast.

For the change in the degree of lip spread for /ɑ/ and /ɔ/, "Increase" indicates that there was a significant increase in the degree of lip spread compared to normal speech, while "Decrease" indicates that there was a significant decrease in the degree of lip spread. The pattern predicted above was that the degree of lip spread should decrease for /ɔ/, as a result of increased lip rounding, while the degree of lip spread for /ɑ/ should increase. That is, the lip spread measurements for /ɑ/ and /ɔ/ should diverge. Two speakers, CHI010 and CHI001, exhibit this pattern, while eight speakers exhibit a change in one vowel but not the other. The lip spread columns for these speakers in Table 3.3 are given in boldface type. In total, ten of the fifteen speakers appear to enhance the lip rounding distinction for /ɑ/ and /ɔ/ in careful speech in some respect, supporting the hypothesis that speakers optimize their speech for

visual perceptibility. As will be discussed in Chapter 6, there are several factors that make it difficult to make strong conclusions on the basis of these results, but they nevertheless suggest that future studies of visual cue enhancement are warranted.

Five speakers, however, did not increase the degree of lip spread between /ɑ/ and /ɔ/. This is because they adjusted the degree of lip spread for both vowels in parallel, such that no overall change in lip spread contrast is observed, or they decreased the degree of lip spread for /ɑ/, making it more similar to /ɔ/. For two of these speakers, CHI002 and CHI008, the overall difference between /ɑ/ and /ɔ/ in terms of lip rounding does appear to increase, even though both vowels change significantly in the same direction. Thus, a more fine-grained quantitative classification of changes in lip rounding would likely reveal that these two speakers do in fact enhance the lip rounding contrast between /ɑ/ and /ɔ/. However, such a classification is left for future analyses and for now, these two speakers are conservatively classified as not enhancing the lip rounding contrast. A third speaker, CHI019, exhibits a near-merger of /ɑ/ and /ɔ/, as discussed above; as such, it is not necessarily predicted that she will enhance the COT-CAUGHT contrast at all. While she does produce a small increase in the acoustic contrast between these vowels, driven by a difference in tongue shape, it will be shown in Chapter 4 that her perception of these vowels is not affected by visual lip rounding cues. CHI011 decreases the degree of lip spread of /ɑ/, and despite the fact that this speaker produces a tongue position distinction between these vowels in careful speech, there is an overall decrease in acoustic contrast between the two vowels. Thus, this speaker enhances the COT-CAUGHT contrast neither acoustically nor visually. The results for the last of these speakers, CHI005 are somewhat dubious. While this speaker shows a substantial increase in the acoustic contrast between /ɑ/ and /ɔ/, this finding is belied by the articulatory results. The articulatory data show that this speaker not only produces no distinction between /ɑ/ and /ɔ/ in terms of tongue position, but also that the lip rounding difference between /ɑ/ and /ɔ/ is smaller in careful speech than in normal speech. It is likely that the methods used to

measure lip rounding or tongue position simply do not provide sufficient detail to capture this speaker's articulatory patterns. Further investigation is needed.

In sum, the contrastive speech task shows that nearly all speakers attempt to enhance the acoustic contrast between these two vowels in careful speech, and that, for the majority of these speakers, enhancement involves an increase in the use of visible labial gestures. These findings provide tentative support for a model in which speakers actively optimize their speech patterns not only for auditory perceptibility, but also for visual perceptibility. As will be discussed in Chapter 6, however, additional investigation is ultimately needed to elucidate the articulatory and acoustic strategies used to enhance contrasts involving visible labial gestures.

3.6 Chapter Summary

The results of this chapter have shown that the strength of the acoustic contrast between /ɑ/ and /ɔ/ varies by speaker age, such that older speakers generally exhibit a stronger contrast between the two vowels, while younger speakers exhibit a weaker contrast. Articulatory data show that the speakers with the strongest COT-CAUGHT contrast distinguish the vowels with differences in both tongue position and lip rounding, and that the majority of speakers with a weaker COT-CAUGHT contrast distinguish the vowels through lip rounding alone. Only one speaker was observed to produce a contrast between /ɑ/ and /ɔ/ with a difference in tongue position alone; it is hypothesized that this articulatory configuration is dispreferred due to the absence of visual lip rounding cues. This hypothesis is tested in an audiovisual perception experiment in the following chapter. The results of the contrastive speech task also provide tentative support for the hypothesis that speakers optimize their speech for visual perceptibility, with the majority of speakers enhancing the visible rounding contrast

between /ɑ/ and /ɔ/. The implications of these findings for theories of sound change, as well as limitations of the careful speech task, will be discussed in Chapter 6.

Chapter 4

Audiovisual Speech Perception in the Maintenance of Phonological Contrast

In Chapter 3, it was shown that Chicagoans differ in the degree and manner by which they maintain the COT-CAUGHT contrast. While the majority of speakers were shown to distinguish /ɔ/ from /ɑ/ with a difference in both tongue position and lip rounding, some speakers maintain the contrast with a difference in lip rounding alone, at least in a normal speech context. A third pattern, in which speakers maintain the contrast solely with a difference in tongue position, was observed only for a single speaker, suggesting that this articulatory configuration may be rare or dispreferred. The rarity of this pattern, however, is not predicted by phonetic models where the speaker's output target is defined in terms of acoustics. These types of models find support from articulatory variation observed for sounds like /ɹ/, where the acoustic target is a low F3, or /s/, where the acoustic target is high-frequency, high-amplitude noise. For these sounds, it appears that the articulatory configuration a speaker chooses is generally arbitrary, as long as the appropriate acoustic output is achieved. These sounds differ crucially from multiply-articulated segments (such as /ɔ/ or /u/), however, in that the articulatory variation observed for /ɹ/ and /s/ generally involves variability in tongue shape, rather than variability involving more visible articulators, such as the lips.[1]

As discussed in Chapter 1, a wealth of evidence from studies on non-auditory speech perception has shown that perceivers of language are sensitive to a wide range of perceptual

1. Labialized variants of /ɹ/, in which /ɹ/ is realized as a labiodental approximant [ʋ], also occur in English (see Foulkes and Docherty 2000), but these variants differ from other /ɹ/ variants acoustically as well as articulatorily. As discussed in Chapter 1, tongue shape variation for /ɹ/ is generally considered to be auditorily indistinguishable, at least in American English.

modalities, including not only auditory cues, but also visual and tactile cues. Thus, one plausible explanation for the rarity of unround variants of /ɔ/ is that visual lip rounding cues play a central role in maintaining a perceptual contrast between /ɑ/ and /ɔ/, even when the acoustic difference between these vowels is diminished. This hypothesis was previously tested by Havenhill and Do (2018), who investigated the audiovisual perception of the /ɑ/-/ɔ/ contrast among Michiganders. They find that the absence of visual lip rounding cues for /ɔ/ causes perceivers to perceive an auditory /ɔ/ stimulus as /ɑ/. This finding suggests that visible lip rounding cues may aid in discrimination of /ɑ/ and /ɔ/, making articulatory variants where /ɔ/ is produced with unround lips perceptually weaker than those where /ɔ/ retains its rounding.

Recent studies have addressed the synchronic relevance of visual lip rounding cues to vowel perception (Traunmüller and Öhrström 2007a, 2007b), as well as the potential role of visual speech perception cues in diachronic sound change (Johnson, DiCanio, and MacKenzie 2007; McGuire and Babel 2012; Johnson 2015). However, the role of visual perception of lip rounding has not previously been considered with respect to its implications for articulatory variability. This chapter presents results from an experiment designed to investigate the audiovisual perception of /ɑ/ and /ɔ/ among perceivers from Chicago, and the relationship between perceivers' own production of lip rounding for /ɔ/ and their perceptual behavior.

4.1 This Experiment

The primary research question this experiment seeks to address is: what is the role of visual lip rounding cues in maintaining perceptual contrast between /ɑ/ and /ɔ/? If visual cues are shown to play an important role in listener identification of /ɑ/ and /ɔ/, this would provide an explanation for the rarity of articulatorily unround variants of /ɔ/. As with the production

experiment presented in Chapter 3, however, this chapter also seeks to address outstanding issues raised by Havenhill and Do (2018) in their study of audiovisual speech perception among Michiganders. While Havenhill and Do (2018) found that listeners were more likely to perceive a visually unround /ɔ/ stimulus as /ɑ/, there was a range of individual variation in the degree and direction of misperception of incongruous audiovisual stimuli. They suggest that this variability in perception may be explained by perceivers' own patterns of production. They hypothesize that perceivers who produce /ɔ/ with unround lips in their own speech will exhibit less of an effect of audiovisual incongruity, given that they do not rely on lip rounding to produce the COT-CAUGHT contrast contrast in their own speech. Thus, the second goal of this chapter is to consider more closely the relationship between production and perception, by analyzing perceptual data from the same participants who took part in the production experiment presented in Chapter 3.

4.2 Methods

4.2.1 Participants

The same participants completed the perception experiment as took part in the production experiment. However, one participant, CHI003, was unable to complete the perception experiment due to an error that caused the experiment software to irrecoverably crash.

4.2.2 Materials

The stimulus list for the perception experiment contains 120 items and is presented in Appendix B. Video recordings of 60 monosyllabic nonce words were created, including 10 words for each of the vowels /ɑ ɔ i u e o/. As in the production experiment, the target vowels were /ɑ/ and /ɔ/, while words containing /i u e o/ served as fillers and controls. Nonce words were generated by finding combinations of phonotactically legal onsets and codas which

did not form a real English word when any of the target or filler vowels were inserted, and that rhymed with at least one real English word. Stimuli were produced by talkers raised in Metro Detroit who exhibit the Northern Cities Vowel Shift and who produce /ɔ/ with a measurable degree of lip rounding.[2] Four talkers (2 men, 2 women) were included to control for differences in talker sex on the effect of visual speech speech cue integration, as well as other intertalker differences (Gagné et al. 1994; Kricos 1996; Traunmüller and Öhrström 2007a; McGuire and Babel 2012). As a result, a total of 480 stimuli were created.

For three of the four talkers, stimuli were recorded in a sound-attenuated booth at Georgetown University; stimuli from the fourth talker were recorded in a quiet room in the talker's home. In both locations, stimuli were filmed in front of a plain backdrop. Video was recorded at a resolution of 1920 × 1080 pixels at 120 fps using a Sony RX10-III digital camera. Audio was simultaneously recorded with an AKG C-417L lavaliere microphone and a Focusrite Scarlett 2i2 USB audio interface. The talkers' rate of speech was controlled by asking talkers to time their utterances to a timer bar displayed in PsychoPy (Peirce 2007). Audio for each nonce word was extracted using Praat. Pink noise was added at a −15 dB signal-to-noise ratio and the mean amplitude of each stimulus was scaled to 70 dB.[3] Each audio recording was then paired with one of two video recordings: the original, congruous video, and video that was incongruous in lip rounding (for target items) or in height (for control items). For target items, recordings of words containing /ɑ/ and /ɔ/ were mismatched to produce round and unround variants of these vowels. For control items, recordings of the vowel pairs /i e/ and /u o/ were mismatched to produce visually high

2. Talkers were chosen from among the participants of the production experiment conducted by Havenhill and Do (2018). All four talkers were therefore known to produce /ɔ/ with visibly round lips. One of the talkers also served as the talker for the perception experiment presented there, but a new set of stimuli were recorded.

3. Addition of pink noise and amplitude scaling were accomplished using a Praat script based on Daniel McCloy's "Mix Speech With Noise" script available at https://github.com/drammock/praat-semiauto. Pink noise was generated using the formula provided in David Weenink's (2014) *Speech Signal Processing with Praat*.

and mid variants of each vowel. Video editing was performed using the command line tool `ffmpeg` (FFmpeg Developers 2018). When necessary, duration of the video was scaled (on a segment-by-segment basis) to match that of the incongruent audio. The `minterpolate` filter was used to smoothly interpolate between frames when increasing the duration of the video. Video for each talker was cropped such that the apparent size of the talker's head was consistent across talkers, and the position of the talker's mouth was centered at the lower third of the video. After editing was complete, the stimuli were downsampled to a resolution of 1280 × 720 pixels and a frame rate of 60 fps for presentation.

Prior to running the experiment, stimuli were verified for naturalness by two independent raters who were naïve to the purpose of the experiment. Stimuli were presented to each rater in pseudo-random order, with the stimuli for each talker presented in separate blocks. For each stimulus, raters were asked to indicate whether the audio and video were synchronized by pressing a button labeled "okay" or "not okay." A third option of "other" was also provided to identify any other issues not relating to audio-video synchronization, such as excessive blinking or non-neutral facial expressions. Stimuli that received a rating of "not okay" or "other" from one or both raters were checked for issues and, when necessary, manually re-aligned or replaced with video from another take. The new set of stimuli was then re-checked in the same manner by a single rater, after which none of the stimuli were identified as having issues.

4.2.3 Procedure

As with the production experiment described in Chapter 3, data for the perception experiment were collected at Northwestern University in Evanston, Illinois. Perception and production data were collected in the same session, with a short break between tasks. The perception experiment was run after the production experiment, in order to avoid influence from the talkers' speech on participants' production patterns.

Participants were seated in a sound-attenuated booth approximately one meter away from a 27 inch computer monitor, with video presented approximately at eye level. Audio was presented to participants through AKG K701 headphones. Stimuli were presented in pseudo-random order with PsychoPy (Peirce 2007). As in the production experiment in Chapter 3, the randomized stimulus list was generated in such a way that no two stimuli containing the same vowel were presented in successive order, nor were two stimuli containing both members of a vowel pair (i.e., /e/ stimuli were not followed by /i/, /o/ stimuli were not followed by /u/, /ɑ/ stimuli were not followed by /ɔ/, and vice versa). The stimuli for each talker were presented in separate blocks, with block order randomized by participant. Participants were given the opportunity to take a break of up to one minute between blocks.

Participants were instructed to observe each stimulus and indicate which existing English word rhymed with the word spoken by the talker.[4] After a stimulus was presented, participants identified the perceived vowel by selecting a rhyming word of English from one of two choices presented on screen. A 2000 millisecond time limit was imposed on responses, after which the program automatically advanced to the next stimulus. The experimental design is presented schematically in Figure 4.1. Participants selected their response by pressing a colored button on a Cedrus RB-30 response pad, which recorded both their response and their reaction time (calculated using the response pad's internal timer). Participants were given five practice trials (using real words rather than nonce words) at the beginning of the experiment.

The rhyming task was chosen in place of a more typical identification task due to ambiguities in English orthography, particularly for /ɑ/ and /ɔ/. In similar audiovisual perception experiments, participants have been asked to identify the percept either by writing down the (nonce) word they heard (Traunmüller and Öhrström 2007a) or by pressing a key on a keyboard labeled with the corresponding vowel (Harrington, Kleber, and Reubold

4. The full set of instructions for the perception experiment is given in Appendix B.

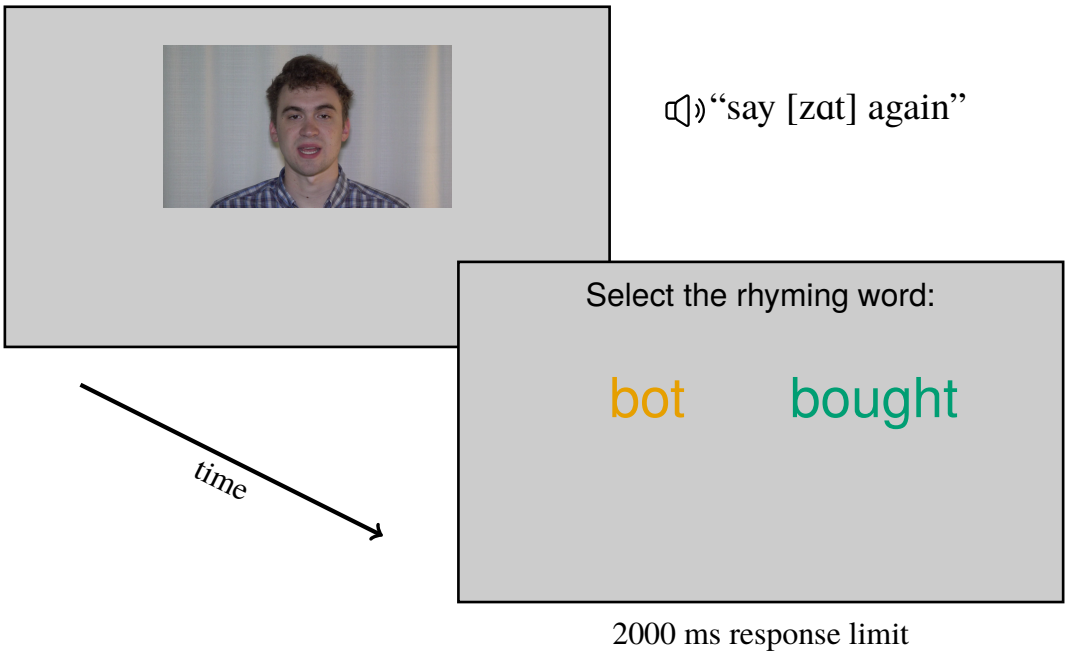


Figure 4.1: Perception experiment design.

2011). However, these particular studies investigated perception among speakers of Swedish and German, respectively, both of which have more regular phoneme-grapheme correspondences than English. In this experiment, participants were asked to identify the vowels /ɑ/ and /ɔ/, which exhibit overlap in their English orthographic representations: /ɑ/ can be represented as ⟨o⟩, ⟨a⟩, or ⟨ah⟩, while /ɔ/ is represented by ⟨o⟩, ⟨a⟩, ⟨au⟩, ⟨ough⟩, ⟨augh⟩, and ⟨aw⟩. Thus, it would be difficult to determine which vowel a participant perceived if the task relied on asking participants were asked to press a labeled button or write down the percept in a pseudo-orthography.

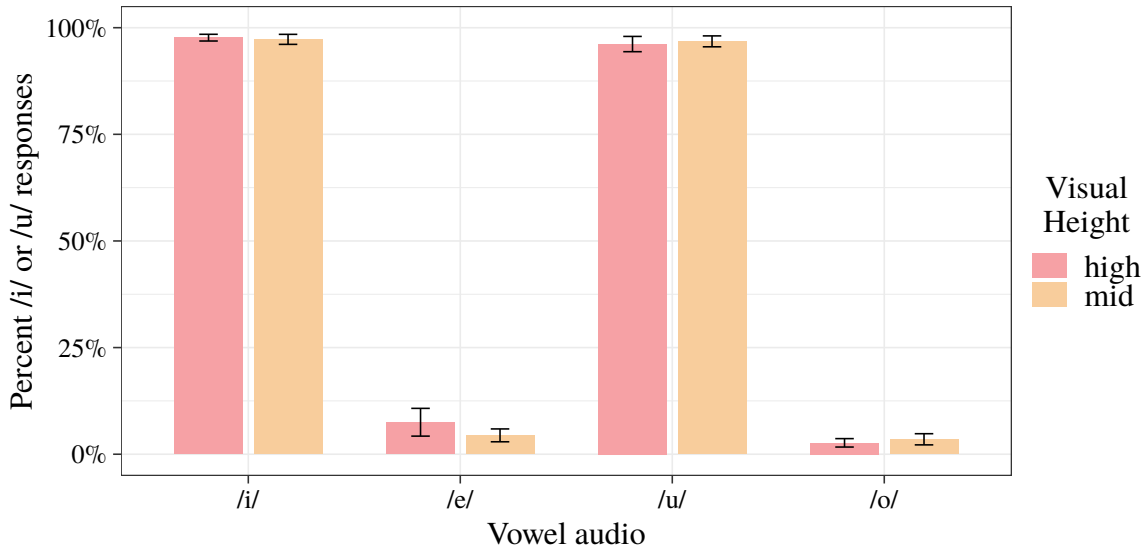


Figure 4.2: Perception results for control items, all participants. Error bars indicate standard error.

4.3 Results

Figure 4.2 shows perception results for the control stimuli, /i e u o/, which were visually incongruous in terms of vowel height. For the high vowels, /i/ and /u/, participants identified the stimulus as high in 97% of trials. A two-sample t-test run for each vowel shows that there is no significant difference between visually high and visually mid stimuli for either /i/ (t(20.92) = 0.508, p = 0.617) or /u/ (t(26.338) = -0.151, p = 0.881). /e/ was identified as high in 6% of trials, which is somewhat higher than expected.[5] However, as with the high vowels, there was no significant effect of visual incongruity (t(19.92) = 0.831, p = 0.416). Finally, /o/ was identified as high in 3.1% of trials, with no significant difference between

5. This figure appears to be driven by a single participant, CHI015, who identified auditorily mid/visually high tokens of /e/ as /i/ in 37% of trials.

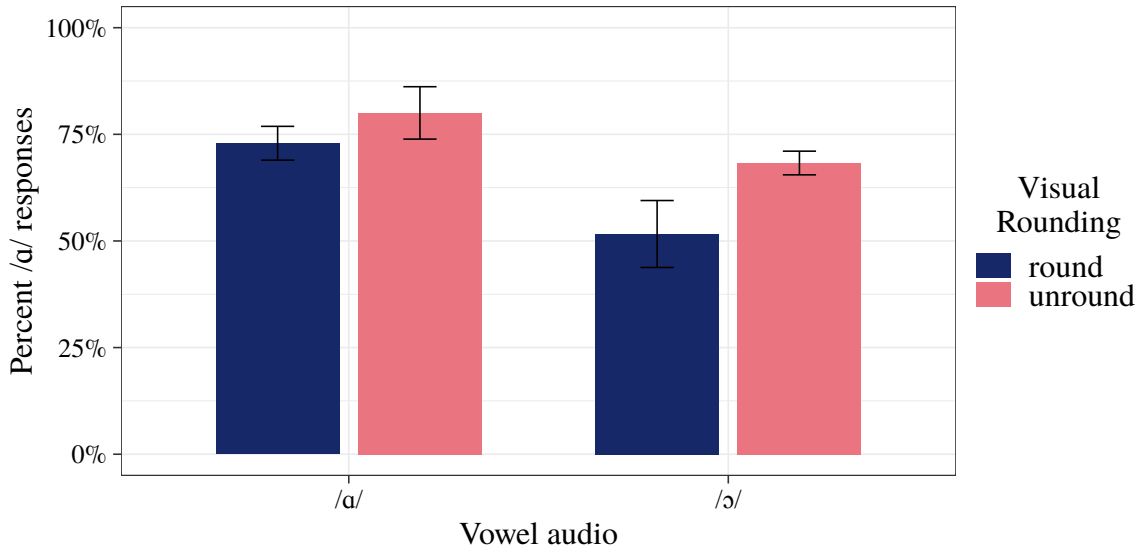


Figure 4.3: Perception results for perceivers (*N* = 7) who distinguish /ɔ/ from /ɑ/ with both lip rounding and tongue position. Error bars indicate standard error.

congruous and incongruous stimuli (t(26.065) = -0.508, p = 0.616). In sum, audiovisual incongruity did not have a significant effect on participants' perception of vowel height.

A significant effect of visual incongruity is observed, however, for the target items /ɑ/ and /ɔ/, which were mismatched in terms of lip rounding. Figure 4.3 presents perception results for participants who contrasted /ɔ/ from /ɑ/ with both lip rounding and tongue position in the production experiment. For these participants, auditory /ɑ/ was correctly identified as /ɑ/ when presented with unround lips in 80% of trials. Perception of auditory /ɔ/ was substantially less accurate; participants in this group identified auditory /ɔ/ as /ɑ/ in 51.6% of trials, roughly at chance, even when presented with congruous lip rounding. Thus, listeners appear to exhibit a bias toward hearing /ɑ/, even when the stimulus presented is /ɔ/. One potential explanation for this bias is frequency—Mines, Hanson, and Shoup (1978),

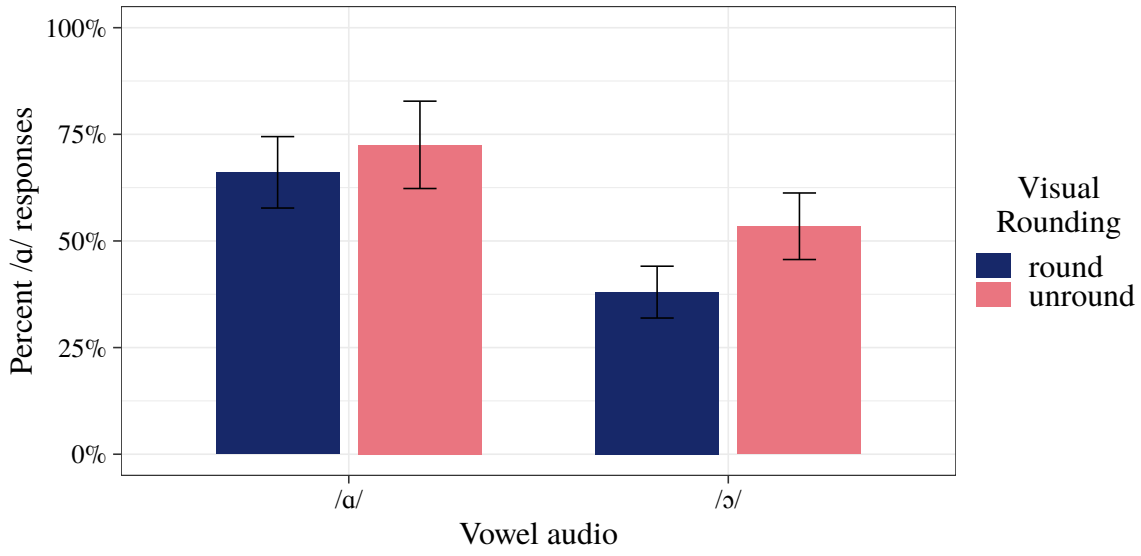


Figure 4.4: Perception results for perceivers (*N* = 6) who distinguish /ɔ/ from /ɑ/ with lip rounding alone. Error bars indicate standard error.

for instance, show that the token frequency of /ɑ/ in conversational American English is approximately twice that of /ɔ/. When paired with incongruous video of unround lips, however, auditory /ɔ/ was perceived as /ɑ/ in 68.3% of trials. An effect of incongruity was not observed for /ɑ/: 72.9% of auditory /ɑ/ stimuli were perceived as /ɑ/ even when presented with visual lip rounding cues.

A similar pattern is observed for participants who distinguished /ɔ/ from /ɑ/ through lip rounding alone, as shown in Figure 4.4. For these participants, identification of /ɔ/ was somewhat more accurate; auditory /ɔ/ was identified as /ɑ/ in 38% of visually congruous trials. When presented with unround lips, however, /ɔ/ was identified as /ɑ/ in 53.4% of the time. Again, there was no effect of visual incongruity for auditory /ɑ/ stimuli, which

Table 4.1: Mixed effects logistic regression model for perceivers (*N* = 13) who produce /ɔ/ with lip rounding.

Predictor	Estimate	SE	z value	Pr(>|z|)	
Intercept (/ɑ/, Congruous)	0.759	0.067	11.282	$p < 0.001$	***
Vowel Audio					
/ɔ/	-0.309	0.065	-4.786	$p < 0.001$	***
Visual Congruity					
Incongruous	-0.062	0.065	-0.952	$p > 0.05$	
Audio * Congruity					
/ɔ/ * Incongruous	0.220	0.091	2.406	$p < 0.05$	*

were correctly identified as /ɑ/ in 72.5% of visually congruous trials and 66.1% of visually incongruous trials, a difference which is not statistically significant.

In order to verify the effect of incongruous lip rounding on vowel identification, a mixed effects logistic regression model was built for all perceivers who contrast /ɔ/ from /ɑ/ via lip rounding in their own speech production, i.e., all thirteen perceivers in the two groups above. A summary of the model is presented in Table 4.1. The model was built with fixed effects of auditory vowel class (baseline = /ɑ/), visual incongruity (baseline = congruous), and the interaction of these effects. Random effects of perceiver, talker, item, and presentation order were also included in the model. The significant negative effect of auditory vowel class indicates that stimuli containing auditory /ɔ/ are less likely to be perceived as /ɑ/ than stimuli containing auditory /ɑ/. This result demonstrates that these perceivers are able to reliably distinguish between the two vowels, which is unsurprising given that the model was run only for perceivers who produce a contrast between /ɑ/ and /ɔ/ in their own speech. More interestingly for the present study is the finding that the main effect of visual incongruity is not significant, but that the interaction of visual incongruity and vowel class is significantly

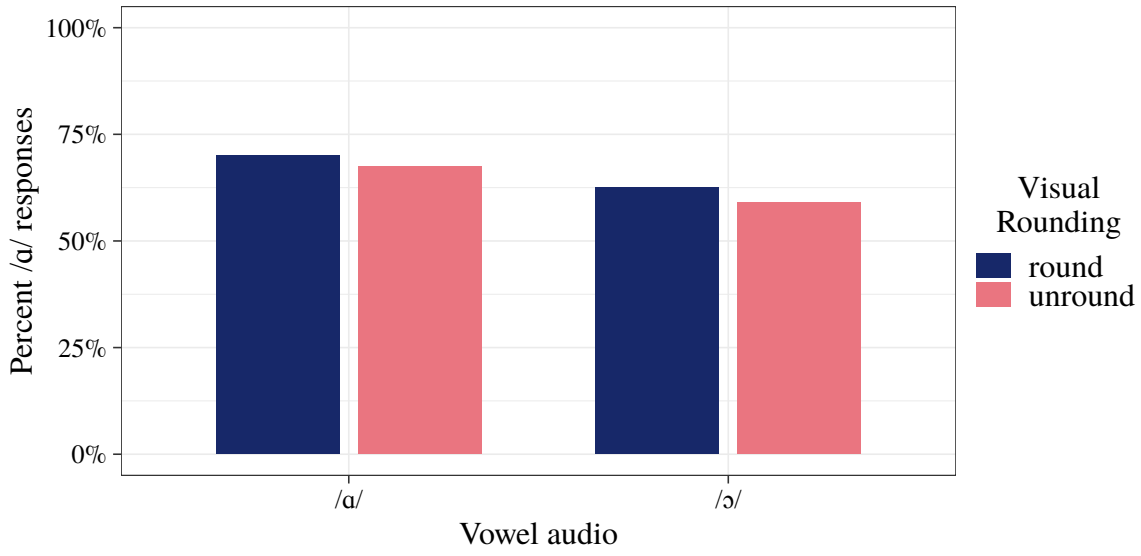


Figure 4.5: Perception results for perceiver (*N* = 1) who distinguishes /ɔ/ from /ɑ/ with tongue position alone.

positive. This result supports the observation seen in Figures 4.3 and 4.4 that /ɑ/ is not perceived as /ɔ/ when produced with visually round lips, but that /ɔ/ is perceived as /ɑ/ when produced with unround lips.

Perception results for the participant who produces the COT-CAUGHT contrast contrast with tongue position alone are presented in Figure 4.5. This participant does not exhibit the same pattern of perception as the previous two groups. While this participant correctly perceives congruous /ɑ/ stimuli in 67.5% of trials, a rate close to that of the other groups, she shows no effect of visual congruity for /ɔ/, which is perceived as /ɑ/ in 62.5% of congruous trials and 59% of incongruous trials. In other words, this participant does not seem to rely on visual lip rounding cues to identify /ɔ/. While it is impossible to make generalizations on the basis of data from a single participant, the prediction stated in Section 4.1 is tentatively

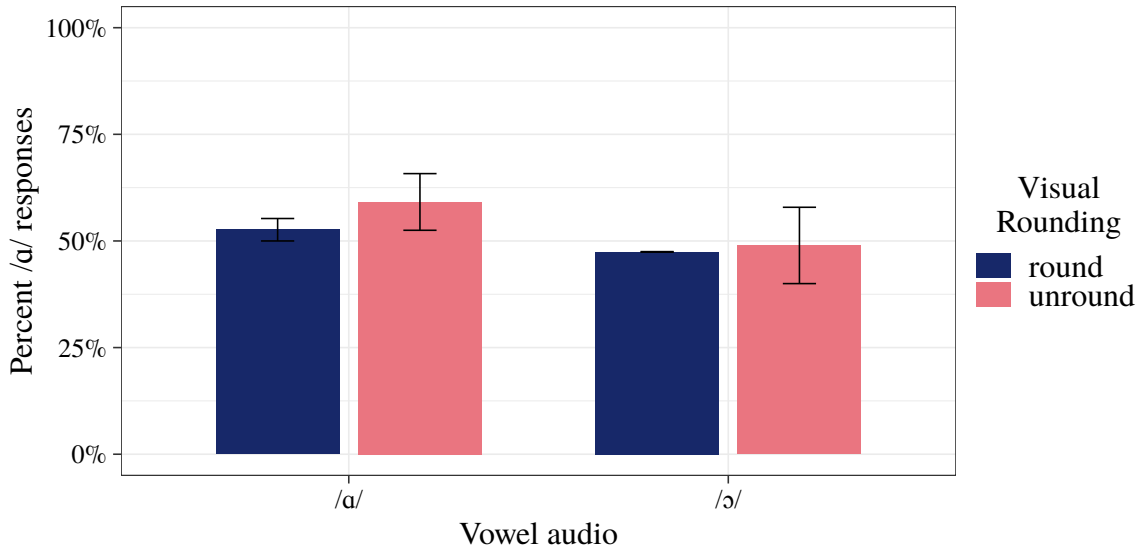


Figure 4.6: Perception results for perceivers (*N* = 2) who do not produce a contrast between /ɑ/ and /ɔ/. Error bars indicate standard error.

borne out: perceivers who do not rely on lip rounding in their own production of /ɔ/ show no effect of visual incongruity, such that /ɔ/ is perceived as /ɑ/ at comparable rates, regardless of whether visual lip rounding cues are present.

Finally, Figure 4.6 presents results for the two participants who did not produce a contrast between /ɑ/ and /ɔ/ in the production experiment. It is observed that these participants correctly perceived /ɑ/ and /ɔ/ at chance, regardless of whether the stimuli were presented with congruous or incongruous lip rounding. This result suggests that these participants have collapsed the contrast between /ɑ/ and /ɔ/ in both production and perception.

4.4 Chapter Summary

The results of the perception experiment presented in this chapter indicate that visual lip rounding cues play in important role in perceiver identification of /ɑ/ and /ɔ/, at least for participants who produce the COT-CAUGHT contrast with a lip rounding distinction in their own speech. For these speakers, stimuli containing auditory /ɔ/ were significantly more likely to be perceived as /ɑ/ when presented with unround lips. This finding has implications for patterns of both articulatory variation and diachronic sound change. In Chapter 3, it was noted that speakers in the Inland North have multiple options for achieving the increase in F2 that is associated with the fronting of /ɔ/, including both tongue fronting and lip unrounding. The results of the production experiment presented there, however, found that a large majority of speakers retain the lip rounding distinction between /ɔ/ and /ɑ/, even if the tongue position distinction is lost. Moreover, speakers were observed to increase the difference between /ɑ/ and /ɔ/ in terms of lip rounding when contrasting the two vowels in careful speech. The results from the perception experiment account for these findings by demonstrating that strategies in which /ɔ/ loses its lip rounding result in a weaker perceptual contrast with /ɑ/. Thus, speakers are predicted to prefer articulations that retain lip rounding because it provides the greatest degree of perceptual contrast across both the auditory and visual domains.

In terms of diachronic sound change, these results suggest that sounds that are visually salient are less likely to be misperceived than sounds that are visually indistinct. One point to be addressed, however, is the question of why some Chicagoans already exhibit a merger of /ɑ/ and /ɔ/, if visual lip rounding cues do in fact inhibit the misperception of these two vowels. Indeed, these vowels are merged throughout much of North America (Labov, Ash, and Boberg 2006). In Chapter 3, it was noted that a number of recent studies have observed a reversal of the Northern Cities Shift throughout the Inland North (Dinkin 2009; McCarthy 2010; Friedman 2014; Driscoll and Lape 2015; Wagner et al. 2016). This reversal is charac-

terized by the lowering of /æ/ and by the backing of /ɑ/ (among other changes), both of which were observed among the youngest speakers in the present study. Havenhill and Do (2018) point out that as /ɑ/ undergoes backing, visual rounding cues for /ɔ/ may inhibit merger of these two vowels, a prediction which is not borne out for the speakers in this study. One explanation for this may lie in the fact that the reversal of the NCS appears to be a change imported from the "third dialect" of American English, rather than a change internal to the NCS itself. This can be understood through the mechanisms proposed by Labov (2007), who distinguishes between two distinct types of sound change: transmission, in which dialect features are passed down from parent to child within a speech community, and diffusion, in which sound patterns spread piecemeal from one variety to another. If reversal of the NCS is the result of diffusion, rather than transmission, there may be pressure to lose, rather than to retain, the rounding contrast between /ɑ/ and /ɔ/. Social pressure to adopt changes associated with the third dialect apparently outweigh system-internal pressures toward retaining the lip rounding contrast.

Another confounding factor is the fact that lip rounding for /ɔ/ may not always be visually salient, particularly when /ɔ/ appears next to another labial segment, such as /p/ (*paw*) or /w/ (*walk*). This is particularly true for low round vowels like /ɔ/, which typically exhibit a smaller degree of rounding than high vowels due to the openness of the jaw (Ladefoged and Maddieson 1996). As will be discussed in the next chapter, there are a number of documented sound changes in which a sound can lose its labial gesture in the presence of a neighboring labial segment, which can be analyzed as misinterpretation on the part of the listener. The issues of transmission and diffusion, as well as variation in the visual salience of lip rounding, demonstrate that there are a multitude of factors involved in language variation and change, and any single factor (such as visual lip rounding cues) can not account for every eventuality. Nevertheless, the experiment presented here provides strong support for the hypothesis that lip unrounding is disfavored as an articulatory strategy for vowel

fronting due to the loss of visual speech cues. The following chapter will explore in more detail historical sound changes in which consideration of visual speech cues can improve our understanding of the mechanisms of language change.

Chapter 5

Labial Persistence in Diachronic Sound Change

The findings presented in Chapter 4 demonstrate that visual speech cues, particularly visual cues to lip rounding, aid perceivers in distinguishing between acoustically similar or ambiguous sounds. In the case of the COT-CAUGHT contrast, identification of auditory [ɔ] stimuli as /ɔ/ was significantly more successful when the acoustic signal was accompanied by video containing visibly round lips. This result suggests that visual lip rounding cues may inhibit misperception of ambiguous sounds, thereby avoiding merger through misperception-based change. Moreover, under a teleological approach to speech production, it predicts that speakers will prefer articulatory strategies that maintain contrast in both the auditory and visual domains. In this case, speakers will prefer to retain lip rounding as vowels undergo fronting, even though unrounding the lips would serve to achieve the same acoustic output. However, given that the experiments presented in this dissertation deal with synchronic perception and production, it cannot be known whether visual cues provide sufficient information to maintain acoustically weak rounding contrasts over time. Indeed, some participants in the study already exhibit a merger of /ɑ/ and /ɔ/, demonstrating the multiplicity of factors that contribute to language change. It may be the case that visual cues are in fact insufficient to preserve phonological contrasts in diachrony, despite their contributions to synchronic perception. Although it is well established that laboratory-based experiments can simulate the conditions that contribute to diachronic sound change (cf. Ohala 1993 and much related work), it is worth considering whether the historical record supports the hypothesis that visual cues can inhibit misperception-based change.

This chapter presents a review of sound changes involving labial segments, including labial-velar alternations, debuccalization, and palatalization of labials. Typological surveys and empirical studies of each type of change suggest that while a loss of labiality is possible, it is generally dispreferred. In cases where segments do lose their labial gesture, that change is most often restricted to environments with a neighboring labial segment. This suggests that learners of the language may have misattributed the source of the labial gesture, but did not fail to perceive it altogether. Thus, changes in labiality seem to predominantly be instances of hypercorrection (in the sense of Ohala 1981, 1993), rather than an absolute loss of labiality. As with the integration of visual speech cues demonstrated in Chapter 4, it is suggested here that visual cues play an important role in enforcing labiality. This is particularly true for the case of palatalization of labials, where labials with secondary palatalization tend to remain labial despite their acoustic similarity to palatals. An account grounded in auditory misperception would predict that labials with secondary palatalization should be susceptible to misperception; instead, visual cues provide robust input to the language learner, helping them to disambiguate the acoustic signal.

5.1 Labial-Velar Alternations

Labialized segments are extremely common throughout the world's languages. 339 (75.17%) of the 451 languages surveyed in the UCLA Phonological Segment Inventory Database (UPSID, Maddieson 1984; Maddieson and Precoda 1990) count a labiovelar glide in their inventories, while 84 (18.63%) contain segments with secondary labialization, making labialization the second most common type of secondary articulation (following nasalization).

Ohala and Lorentz (1977) discuss the cross-linguistic patterning of labiovelars, including the consonants [ʍ, k͡p, g͡b, k^w, g^w], the vowel [u], and the labiovelar glide [w]. Their account is concerned with addressing the question of whether labiovelar segments are primarily

[Labial] or [Dorsal], which they contend is a "pseudo-problem." Ohala and Lorentz argue against "pigeonhole-filling" approaches, such as that of Anderson (1976), where the stops /k͡p/ and /g͡b/ are classified either as labial or velar based depending on gaps in a language's stop inventory. Under this type of approach, labiovelars are classified as labial in languages lacking /p/, but are classified as velar in languages lacking /g/. Chomsky and Halle (1968) similarly argue that labiovelars must either be velars with secondary labialization, or labials with secondary velarization, and cannot be both [Labial] and [Dorsal]. However, Ohala and Lorentz point out that such approaches are problematic not only for languages which have labiovelar stops despite having no gaps in their stop inventory, but also for languages in which labiovelar stops pattern with velars, despite inventory gaps suggesting they be classified as labial. They argue instead that labiovelars are better classified by their phonetic properties, which can cause them to behave either as labial or as velar, depending on the particular context.

Ohala and Lorentz observe a number of cross-linguistic tendencies shared by labiovelars, the first of which is that [w] exhibits properties of both labial and velar segments, and patterns variably with labial or velar obstruents (or both). In addition, there is evidence that labial offglides have developed historically from both labial and velar segments, as in Indo-European, as well as observations that labialization appears as a secondary articulation most frequently on velar, uvular, and labial segments, but not as frequently on coronal consonants. Ohala and Lorentz argue that these patterns are the result of the lowering of F2, a property shared by both velars and labials. This is due to the fact that constrictions at the velum and lips correspond to velocity maxima of the standing wave formed in the vocal tract by the second resonant frequency. They suggest that labiovelars are unique in exhibiting this property, as no other pairs of simultaneous constrictions exert such a strong effect on F2. This property of labial and velar segments has long been observed in phonology, and has

been represented with the feature [grave], defined for segments in which "the lower side of the spectrum predominates" (Jakobson, Fant, and Halle 1951).

Although labiovelars share properties with and can arise historically from both velar and labial segments, Ohala and Lorentz observe that "When becoming a fricative or determining the place of articulation out of adjacent fricatives by assimilation, [w] shows itself primarily as a labial, less often as a velar" (587). They note that this pattern can be found in many languages, citing evidence from at least 20 languages including Javanese, Kirghiz, Hungarian, Yolax Chinantec, Slave, and more. They argue that this tendency originates in the effects of a labial closure on the turbulent noise produced at the velar closure, such that the velar noise is both dampened and modified by the labial constriction. This generalization holds in sound change as well; they note that "Labiovelar obstruents will most likely change to labial not velar obstruents" (589). This tendency has been observed in many sound changes, as shown in (6):

(6) Substitution of labialzed velars with labials (Ohala 1993, 242):

a.	*Indo-European*	>	*Classical Greek*	
	*e**kw**ōs		hi**pp**os	'horse'
	*$\mathbf{g^w}$iwos		**b**ios	'life'
	*ye$\mathbf{k^w}$r̥		he**p**atos	'liver'
b.	*Proto-Bantu*	>	*West Teke*	
	*-**k**umu		**pf**uma	'chief'
c.	*Proto-Yuman*	>	Yuma	
	*imali$\mathbf{k^w}$i		mal^{y}**p**u	'navel'
d.	*Proto-Muskogean*	>	*Choctaw*	
	*$\mathbf{k^w}$ihi		**b**ihi	'mulberry'
	*uN$\mathbf{k^w}$i		um**b**i	'pawpaw'

e. *Sungkhla*

khwài	~	**f**ài	‘fire’
khón	~	**f**ón	‘rain’

f.

Proto-Zapotec	>	*Isthmus Zapotec*	
*$\mathbf{kk^w}$a-		**p**a	‘where’

They cite two counterexamples to this generalization, noting that [u], [w], and [ɣ^{w}] are in free variation in Araucanian, and that Danish exhibits variation between the segments [u], [uu̯], [ux], and [uk]. Interestingly, when the velar variants of these segments appear, labialization is retained due either to the presence of a neighboring [u] or to a secondary labial articulation. These examples may be analyzed as a reinterpretation on the part of the learner in which the labiality of the labiovelar was attributed to a neighboring segment rather than to the labiovelar itself. Under this analysis, these alternations are not only *not* counterexamples to their generalization, but instead provide support for a tendency to preserve labiality.

A tendency for labial and velar segments to alternate only near a neighboring labial segment can be observed in a number of other languages. For instance, Mazzaro (2010) provides an analysis of a synchronic alternation between [ß] and [ɣ] that appears in a number of Spanish dialects. The alternation occurs primarily before the vowel [u] and the diphthongs [we] and [wi], giving rise to realizations of words like *abuelo* ‘grandfather’ as [aɣwelo], rather than [aßwelo]. Data for Mazzaro’s study come from sociolinguistic fieldwork in Caá Catí, Argentina, which was chosen in part for its high rates of illiteracy and low social and geographic mobility, in order to investigate the hypothesis that literacy inhibits labial-velar alternations. This hypothesis is borne out, with literate speakers producing standard [ß] in approximately 84% of tokens, while illiterate speakers produce standard [ß] in only 70% of tokens, otherwise realizing the fricative as [ɣ].

Hickey (1984) discusses labial-velar shifts in a variety of European languages, finding several cases where labials become velars in the presence of a neighboring labial segment.

For instance, he discusses the /k/ > /p/ shift that occurred in the development of Rumanian from Latin. Examples of this shift (from Hickey 1984) are given in (7):

(7)	Latin	Rumanian	
	*co**c**tum*	*co**p**t*	'cooked'
	nox, *no**c**tis*	*nea**p**te*	'night'
	*ac**q**ua*	*a**p**a*	'water'
	***q**uattuor*	***p**atru*	'four'
	*lin**g**ua*	*lim**b**a*	'language'

Hickey finds that much of the previous literature on this shift has dealt with the nature of Latin *ct* sequences, and whether *c* may have represented /k^{w}/. He argues against this view and suggests that it is difficult to see how this would have affected words like *coctum* and *noctis*. However, it seems plausible that /kt/ in *coctum* would exhibit rounding due to both anticipatory and perseverative coarticulation from the surrounding round vowels. While /kt/ in *noctis* would only exhibit perseverative coarticulation from the preceding /o/, it is possible that the pattern of /k/ > /p/ could have been extended to this environment, or that perseverative coarticulation is sufficient for the change to occur. This seems to have been the case in Albanian, where /kt/ became /it/ after front vowels, but /ks/ became /fs/ after back vowels. This pattern is observed in (8), from Hickey (1984):

(8)		Latin	Albanian	
	a.	*dire**ct**us*	*dre**it**ë*	'direct'
		*tra**ct**are*	*tra**it**oj*	'prepare, cook'
	b.	*co**x**a*	*ko**fs**hë*	'hip'
		*la**x**a*	*la**fs**hë*	'battle'

Cahill (1999) describes a number of phonetic and phonological aspects of labial-velar stops and, like Ohala and Lorentz (1977), finds that labial-velar stops may sometimes behave

as velars and sometimes as labials. However, he argues that it is rare for labial-velars to exhibit velar properties, and suggests analyses of labial-velar stops under Feature Geometry and Articulatory Phonology, where either labial or both labial and velar articulations can be given equal prominence. Most relevant for the present purpose, however, is Cahill's discussion of the historical development of labial-velars. He states that "the immediate predecessors of labial-velars seem always to be labialized consonants, whether *Pw or *Kw" (166). This can be observed, for instance, in the Sawabantu group of languages, as shown in (9):

(9)	E. Sawabantu	W. Sawabantu	
	/**kw**alé/, /**kw**adé/	/**kp**aé/	'partridge'
	/**kw**átá/	/**kp**átá/	'sword'
	/**kw**édí/	/**kp**élí/	'worm'

In the Eastern Sawabantu group, the prefix **ku-* is realized as **kw-* before vowels, which developed into *kp* in Western Sawabantu (Mutaka and Ebobissé 1996, cited by Cahill 1999).

Cahill further finds that labial-velar stops tend to change into labial (not velar) consonants, such that [kp] merges into either [gb] or [p]. Both patterns can be found in Senufo languages, where the /k͡p/-/g͡b/ distinction has collapsed to /g͡b/ in Shenara and Sucite, as shown in (10):

(10)	Cebaara	Shenara	Sucite	Supyire	
	/**k͡p**āʔā/	/**g͡b**aʔa/	/**g͡b**āxā/	/**b**āgā/	'house'
	/**g͡b**āʔālāgà/	/**g͡b**aʔalaga/	—	/**b**àhàgà/	'bedbug'

This merger has progressed further in Supyire, where /g͡b/ has merged with /b/, causing /b/ to become disproportionately frequent in the language. Cahill summarizes both types of change, suggesting that the common historical pattern of development for labial-velars is the progression Ku > K^w > KP, where K is any velar stop and P is any labial stop. This

progression may then be extended through a change like k͡p > g͡b and/or KP > P. He argues that this account also explains some synchronic restictions on the distribution of vowels following labial-velar stops, where /u/ and /o/ are the most frequent vowels to be missing after labial-velar stops. This is expected if labial-velar stops are derived from historical /k^{w}/, because labialzed velar stops generally do not occur before round vowels.

The development of labial-velar stops from round vowel + velar sequences is also observed in Vietnamese, one of the few languages outside of Africa to exhibit doubly-articulated labial-velar stops. In the Hanoi variety, word-final velars are realized with rounding (Thompson 1987) or as labial-velar stops (Kirby 2011) when they follow a round vowel, as observed in (11):

(11) Labial-velar final stops in Vietnamese (383):

a.	/su**k͡p**˧˥/	*xúc*	'to scoop'
	/sup˧˥/	*súp*	'soup'
b.	/ho**ŋ͡m**˧˥/	*hông*	'hip'
	/hom˧˥/	*hôm*	'day'
c.	/hɔ**k͡p**˨˩/	*học*	'to study'
	/hɔp˨˩/	*họp*	'to meet'

The origin of these labial-velars as velars is reflected through their orthography, and through the absence of velar stops in this environment. Moreover, Cahill's (1999) observation that labial-velars tend to change into labials is found in Vietnamese as well. Thompson (1987) writes that "this strongly rounded /k/ occasions considerable difficulty for learners of the language, since they will frequently think they are hearing /p/" (26). While a change from labial-velar to labial has not yet occurred in Vietnamese, it is reasonable to suspect that these segments will be susceptible to merger due to their similarity in both the auditory and visual domains.

Evidence suggests, then, that labial-velars tend to originate from velars occurring next to a labial segment, and that they tend to become labials in the course of historical change. While labial-velars can alternate with velars, or become velar themselves, this pattern is observed only in the context of a neighboring labial segment, especially round vowels. In both cases, the labial gesture is retained, with a reduction in duration in the latter case, and at the expense of the velar gesture in the former. Finally, it is possible for velar segments to become labial (ostensibly without an intermediate labial-velar stage) when a neighboring round vowel is present, as observed in the data from Rumanian and Albanian.

Potentially problematic for this generalization, however, is the fact that in certain cases, labiovelars may become velars even in the absence of a neighboring labial segment. Ohala and Lorentz (1977) observe that "When becoming a nasal or determining the place of articulation of adjacent nasals by assimilation, [w] shows itself as a velar, rarely as a labial" (584). They cite data from Kpelle (Welmers 1962), given in (12), in which /w/ patterns with velars in the process of nasal place assimilation and fusion.

(12)

Indefinite	Definite	
/ɓɔ́ɔ/	/ˡmɔ́ɔi/	'wax'
/lúu/	/ˡnúui/	'fog, mist'
/**ɣ**îla/	/ˡ**ŋ**ilaĩ/	'dog'
/**w**ée/	/ˡ**ŋ**wéei/	'white clay'

Ohala and Lorentz argue that this tendency, of which they cite several additional examples, is the result of resonance patterns within the vocal tract, as shown in Figure 5.1. Because all nasal consonants exhibit nasal resonance as a result of the open velum, place differences among nasals are distinguished in part by oral anti-formants. Because nasalized [w̃] has a velar constriction, the oral anti-formants for [w̃] are more similar to [ŋ] than to [m]. Thus, in the case of nasals, there is a strong acoustic motivation for labiovelars not to pattern with labials.

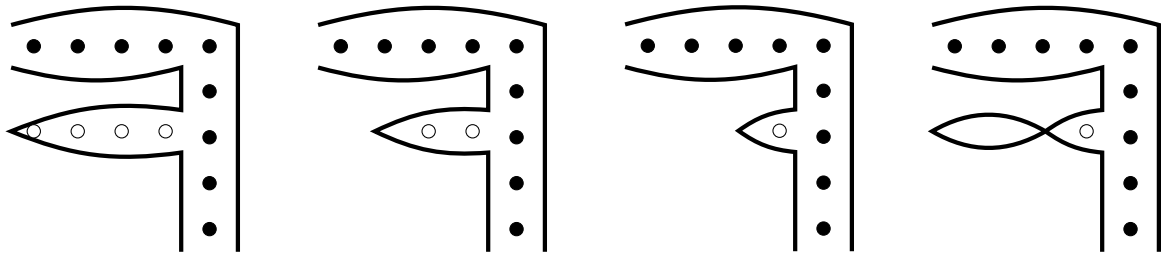


Figure 5.1: Vocal tract resonances for [m], [n], [ŋ], and [w̃]. Filled circles indicate formants, open circles indicate anti-formants. Adapted from Ohala and Lorentz (1977).

Finally, Ohala and Lorentz observe that [w] has a tendency to remain labial when assimilating to adjacent vowels. They have found that in at least 30 languages, /w/ is realized as [w] before back (round) vowels, but appears as [v] or [ß] before front vowels. Their explanation for this pattern is articulatory: in cases where the lingual articulation for /w/ assimilates to a neighboring vowel, the velar constriction remains only near vowels which themselves have a velar constriction, i.e., back vowels. Near front vowels, the velar constriction is lost in favor of the fronted tongue position of the vowel, but given the independence of the lips, the labial articulation remains.

5.2 Debuccalization

In the case of labial-velar alternations, it was observed that labial segments tend to remain labial, except where their labial gesture can be attributed to a neighboring segment. Another natural testing ground for this hypothesis is debuccalization, in which a segment completely

loses its oral articulation and is realized with a laryngeal place of articulation. This type of change is illustrated by changes in the English *wh* series, presented in (13):[1]

(13)	Old English		Modern English	
	/hwaː/	>	/huː/	who
	/hwæt/	>	/wʌt/	what
	/hwæːr/	>	/wɛɹ/	where
	/hwannə/	>	/wɛn/	when
	/hwyː/	>	/waɪ/	why

In most varieties of Modern English (Wells 1982, 228-230; Labov, Ash, and Boberg 2006, 49-50), historical /hw/[2] has merged with the labiovelar approximant /w/, as observed in the interrogatives *what*, *where*, *when*, and *why*. This merger is also observed in homophone pairs like *whine* and *wine*, *whale* and *wale*, etc. In cases like *who*, however, /hw/ is debuccalized to /h/, a difference which can be attributed to the presence of the following round vowel /u/. Where /hw/ precedes a round vowel, the lip rounding present on /hw/ may be interpreted by listeners as anticipatory coarticulation associated with the vowel, rather than as a primary feature of the consonant. This type of change is an instance of hypercorrection (in the sense of Ohala 1993), in which listeners mistakenly analyze as coarticulation and correct for some part of the speech signal that was in fact intended by the speaker. In the case of *who*, rounding produced intentionally for /hw/ would have been analyzed as a coarticulatory effect of the following /u/, leading listeners to postulate that [hw] was an unintended surface variant of /h/. No such misinterpretation is possible when /hw/ precedes an unround vowel, so rounding remains associated with the consonant.

1. The Old English forms in (13) and the Germanic reconstructions in (14) were compiled from Watkins (2000), P. S. Baker (2012), and *OED Online* (2016)

2. /hw/ is variably analyzed as a fricative ([ʍ]; Wells 1982, 228), as a voiceless approximant ([w̥] or [ʍ]; Ladefoged and Maddieson 1996, 326), or as a [h + w] cluster, which was its original pronunciation in Old English (Minkova 2004).

Such a change has occurred several times throughout the history of English, as observed through the orthographic difference between *who* and *how*. In Modern English, both *who* and *how* are pronounced with initial /h/, and both are reflexes of Germanic **hw-*, which itself derives from Proto-Indo-European **k^w-* (Watkins 2000). *How* is the sole member of the *wh* series to be spelled with initial ⟨h⟩, however, a reflection of the fact that both *who* and *how* underwent /hw/ > /h/ changes but did so at different stages in the history of the language. These changes are illustrated in (14):

(14)	Germanic		Old English		Modern English	
	*/hwas/	>	/hwaː/	>	/hu/	who
	*/hwoː/	>	/huː/	>	/haʊ/	how

When /hw/ > /h/ occurred between Germanic and Old English, it affected *how* but not *who*. At the time of this change, *how* contained the round vowel /oː/, allowing for the labial gesture of /hw/ to be misinterpreted as part of the vowel. It was not until a later vowel shift caused the vowel in *who* to become round that the environment necessary for /hw/ > /h/ appeared in *who*. While *how* underwent the change prior to the codification of English orthography, *who* did so only after its spelling had been cemented with initial ⟨wh⟩. Like the labial-velar alternations discussed above, debuccalization of labials occurs more readily in the presence of a neighboring labial segment, such as a following round vowel.

This pattern is also observed cross-linguistically. Foulkes (1997) presents a typological survey of diachronic /p/ > /h/ and /f/ > /h/ changes. In a survey of the 317 languages sampled in the UCLA Phonological Segment Inventory Database (UPSID), he finds nine languages that lack /p/ or /b/ and in which a direct /p/ > /h/ change is reported. The language families in which this change is found include Eastern Oceanic, Tupi-Guarani, Dravidian, Salish, Armenian, Uto-Aztec, Aleut, Japanese, and Bantu. Typical of these changes is that observed in Tupi-Guarani, shown in (15):

(15) Tupi-Guarani

*/puku/	→	Asurini	/poko/	‘long’
		Urubu	/puku/	
		Kamayurá	/huku/	
		Sirionó	/hoko/	

The proto-form */puku/ is realized with /p/ in Asurini and Urubu, but with /h/ in Kamayurá and Sirionó. This change is reported by Lemle (1971, cited by Foulkes 1997) to have occurred only before back round vowels and /w/.

Of the nine reported cases of direct /p/ > /h/ changes that Foulkes finds, seven occur in reconstructions, and there are no attested cases of synchronic variation between /p/ and /h/. He suggests it is likely that these apparent /p/ > /h/ changes had an intermediate stage where /p/ first lenited to /f/. This is supported by a /p/ > /h/ change in Japanese for which there does exist written evidence. The Ancient Chinese form */puk/ is written as /hoku/ in Kan’on and Go’on, but cognates in Korean, Hakka, and Fuchow retained the original /p/, as shown in (16):

(16) Sino-Japanese

*/puk/	→	Kan’on	/hoku/
		Go’on	/hoku/
		Korean	/pok/
		Hakka	/puk/
		Fuchow	/pouk/

Citing Karlgren (1915), Foulkes notes that some grammars provide evidence for an intermediate stage in which the words /ha, hi, he, ho/ are written with initial ⟨f⟩.

Unlike direct /p/ > /h/ changes, /f/ > /h/ changes appear to be fairly common and occurred, for instance, in Castilian Spanish, where /h/ was eventually deleted altogether. This change is shown in (17):

(17)

Latin		Spanish	
facere	>	*hacer* /aθer/	‘to do’
filius	>	*hijo* /ixo/	‘son’
fumu	>	*humo* /umo/	‘smoke’

Although /f/ was eventually reintroduced to the Spanish phoneme inventory, Foulkes suggests that it is once again undergoing debuccalization in some South American dialects. He provides an example from a Ricardo Güiraldes novel, where ⟨j⟩ is used to indicate /h/ instead of /f/ in the speech of rural Argentinians, in words like *juerza* ‘strength’, standardly spelled *fuerza*. In other cases, /f/ is lenited to [x] rather than to [h], suggesting that this change is similar to the labiovelar alternation studied by Mazzaro (2010) and described in Section 5.1.

In nearly all cases of /f/ > /h/ change, with no clear counterexamples, the change occurs in the context of a neighboring high back round vowel. This can be observed in the data from Songhai, given in (18):

(18)

Zarma	Kaado	
/**f**u/	/**h**u/	‘house’
/**f**ortu/	/**h**ottu/	‘bitter’
/fiji/	/fiji/	‘to bury’
/fe꞉ji/	/fe꞉ji/	‘ram’
/farga/	/farga/	‘fatigue’

In Kaado, /f/ is realized as [h] before non-low back round vowels, an example typical of the finding that /f/ > /h/ change is apt to arise in the context of a neighboring /u/. In addition to Songhai, this pattern is observed in Spanish, Koiari, Tahitian, Hausa, and South Lappish, among others. Although this change may sometimes spread to additional environments, there are no reported cases of a /f/ > /h/ change which occurs only before /a/ or /i/.

However, because Foulkes (1997) focuses only on debuccalization of labials and does not examine debuccalization of coronals and dorsals, it is not clear whether labials are any more likely to retain their place of articulation than segments at other places of articulation. If labial, coronal, and dorsal segments undergo debuccalization at similar rates, this would clearly call into question the hypothesis that labials tend to retain their labiality. If debuccalization of labials is rare, however, or occurs only as part of a process of debuccalization that also targets non-labial segments, this would provide additional support for the generalization that changes in the place of articulation of labials are avoided. Typological data show that the latter case is borne out. O'Brien (2012) presents a survey of debuccalization processes in approximately 50 languages and dialects. Included in his survey are any alternations which target an oral consonant and result in a laryngeal consonant, including /h/, /ɦ/, and /ʔ/. Of the 50 cases surveyed, 29 result in /h/, 2 result in /ɦ/, and 19 result in a glottal stop. O'Brien finds that dorsal segments are targeted most frequently by debuccalization processes, followed closely by coronal segments. Dorsal segments are targeted in 34 of the debuccalization processes surveyed, while coronal segments are targeted in 30. (The total is greater than 50 because processes targeting multiple places of articulation were counted for each place of articulation.) Ten of the coronal debuccalization processes occur independently of dorsal or labial debuccalization, and of these processes, the most common are /s/ → [h] ('s-aspiration') and /t/ → [ʔ] ('t-glottalization'). Of the 34 processes targeting dorsal segments, 17 occur without simultaneous coronal or labial debuccalization. In contrast, labial debuccalization is much more rare. Labial debuccalization occurs in only 15 of the 50 cases surveyed by O'Brien, and of these 15 cases, only 2 are processes in which labial segments alone are debuccalized. These two cases are the debuccalization of /w/ word-finally and before C in Pipil, and word-initial /p/ → [h] in Kannada. In the vast majority of cases, labial segments are debuccalized only in languages where all stops or

obstruents are debuccalized. While this sample of languages is not necessarily balanced, the overall findings are consistent with other work on debuccalization.

Thus, the data on both debuccalization and labial-velar alternations suggest that while it is possible for segments to altogether lose their labiality, this tends to occur only under certain circumstances. Specifically, labials and labiovelars have been shown to lose their labial gesture when they occur next to round vowels, when an alternation that occurs next to round vowels is extended to non-round vowels, or when there are strong acoustic motivations (as in the case of nasal place assimilation). In the case of the sound changes reviewed so far, there are several explanations for why labial consonants behave differently from coronal and dorsal consonants. First, the lips are independent from the tongue and other articulators, so lip gestures are not affected by other articulatory movements. As will be discussed below, lingual gestures can engage in blending with other lingual gestures, but labial gestures cannot. Moreover, the independence of the lips allows labial gestures to exhibit an extended temporal domain, such that rounding can persist across several segments. This extended temporal domain means that language learners can easily misattribute the source of labiality, thereby giving rise to hypo- or hypercorrection. Third, because the lips are the most anterior point of the vocal tract, labial gestures can simply block any acoustic effects of more posterior gestures. Finally, and most importantly for the present investigation, lip gestures are visually salient, making them relatively easy to perceive in comparison to articulations that take place entirely inside the vocal tract.

The sound changes reviewed so far can potentially be accounted for in terms of any or all of these factors. For instance, the reason that labial-velars tend to become labial rather than velar may be due to the fact that the labial gesture simply blocks the velar gesture, leaving the velar gesture completely or partially inaudible. Visual salience may also contribute to this tendency, because the labial gesture will be both more audible and more visible to the perceiver than the velar gesture. In this case, both the acoustic/auditory and visual factors

make the same prediction: labial-velars will tend to become labial, not velar, in the course of sound change. However, there also exist cases in which the auditory and visual signals provide conflicting information. One such case, discussed in the next section, is palatalization. It is shown that while coronal and dorsal segments freely undergo palatalization, palatalization of labials is typologically extremely rare. This pattern cannot be explained on the basis of auditory perception alone, because the acoustic effects of palatalization are similar for all segments regardless of their place of articulation. However, the visual salience of labials provides an additional cue to their place of articulation, helping listeners resolve the acoustic ambiguity. Considering the role of visual input thus provides a plausible explanation for the fact that full palatalization of labials is exceptionally rare.

5.3 Palatalization

Palatalization encompasses two separate (but related) phonological processes. Secondary palatalization describes the process whereby a segment is modified by a secondary articulation at the palatal or palatoalveolar region. A wide variety of languages exhibit secondary palatalization of stops before a palatal glide or front vowel. This pattern is particularly common in Slavic, as exemplified by the Russian nominative-dative pairs shown in (19):

(19) Secondary Palatalization in Russian (Kochetov 2011):

a. [trap-a]~[trap^{j}-e] ‘path’

b. [s^{j}irat-a]~[s^{j}irat^{j}-e] ‘orphan’

c. [sabak-a]~[sabak^{j}-e] ‘dog’

Full (or primary) palatalization, on the other hand, describes the complete shift of a segment’s place of articulation toward the palate, as observed in the English lexical pairs given in (20):

(20) Full Palatalization in English:

a. *spiri*[t]~*spiri*[t͡ʃ]*ual*

b. *gra*[d]*e*~*gra*[d͡ʒ]*ual*

c. *pre*[s]~*pre*[ʃ]*ure*

d. *plea*[z]*e*~*plea*[ʒ]*ure*

While secondary palatalization is a common phonological process for obstruents produced at any place of articulation, several typological studies (Bhat 1978; Bateman 2007; Kochetov 2011) have shown that full palatalization of labials is exceptionally rare.

Bhat (1978) argues that palatalization can be understood as three separate processes, including tongue fronting, tongue raising, and spirantization. He suggests that each of these processes involves a distinct triggering environment, and all have differing effects on the 'palatalized' segments. Evidence for fronting and raising as separate processes comes from languages in which coronals and velars are affected separately by palatalization. Whereas coronal consonants are affected only by tongue raising, velars may undergo tongue raising, tongue fronting, or both. For instance, Bhat finds that in languages like Eastern Ojibwa, Kashmiri, and English, [j] triggers raising of coronal (but not velar) consonants. In Latin and Hakka, both velars and coronals are raised before [j], but front vowels induce fronting only of velars. The third process, spirantization, may or may not occur along with tongue fronting and raising. In West Slavic, for instance, /t/ and /d/ become affricates before /j/, but in South Slavic, they become palato-alveolar stops rather than affricates (Chomsky and Halle 1968).

With regard to labials, Bhat finds five instances of full palatalization. More frequent, Bhat finds, are cases in which labials undergo secondary palatalization. Secondary palatal articulations on labials can be found in languages as diverse as Japanese, Slavic, Amharic, Irish, Nupe, Western Ossetic, Carib, and others. In other languages, labial segments undergo

spirantization before palatals, which Bhat also considers to be a type of palatalization. This type of alternation occurs in Piro, where /w/ becomes a labial fricative before front vowels, and in Rundi, where 'palatalized' /b/ becomes [v]. It is unclear from Bhat's (1978) survey, however, whether full palatalization of labials is in fact more rare than secondary palatalization or spirantization. Because his survey focuses on distinguishing tongue raising, tongue fronting, and spirantization as distinct types of palatalization, the rate at which these processes affect segments at differing places of articulation is not quantified.

Motivated by the question of the cross-linguistic frequency of palatalization, and its rate of occurrence with respect to segments at each major place of articulation, Bateman (2007, 2011) performed a genetically and geographically balanced survey of palatalization in 117 languages. Bateman's classification distinguishes two types of palatalization: full and secondary. As noted above, full palatalization is characterized by a complete change of a consonant's place of articulation to the palate, such as /k/ → [t͡ʃ], and may or may not include a change in manner (i.e., spirantization). Secondary palatalization includes processes where a consonant receives a secondary palatal articulation, such as /b/ → [bʲ]. Under Bateman's definition (and in contrast to Bhat 1978), spirantization and affrication alone are not considered palatalization, so alternations such as /t/ → [s] and /t/ → [ts] are excluded from her survey.

Of the 117 languages surveyed in Bateman's sample, 58 languages show palatalization and 59 do not. Among the languages exhibiting palatalization, 45 exhibit full palatalization and 32 exhibit secondary palatalization. (Some languages have both, so the total is greater than 58.) Bateman finds that while coronal and dorsal consonants may independently undergo palatalization (full or secondary), labial consonants do not, exhibiting palatalization only in languages where either coronal or dorsal consonants also palatalize. Moreover, there is a sharp contrast between full and secondary labial palatalization in terms of frequency. While secondary labial palatalization is the most frequent type of secondary palatalization,

occurring in 45% of languages with secondary palatalization, full labial palatalization is exceedingly rare, occurring in only 2 of the 45 languages with full palatalization.

The findings of Kochetov (2011) are similar, and provide support for the cross-linguistic rarity of full palatalization of labials. Kochetov's survey includes 64 languages belonging to 17 language families. He classifies palatalization processes into three types, where Type I corresponds to Bateman's (2007) secondary palatalization, Type II corresponds to Bateman's full palatalization, and Type III represents palatalization processes resulting in an anterior coronal, which were excluded from Bateman's survey. Kochetov finds that secondary palatalization (Type I) is common for segments produced at any place of articulation, occurring in 6 of 17 sampled language families. Also common is Type II processes targeting coronal and dorsal segments, as well as Type III processes targeting coronals, with the most common palatalization process overall being /t/ → [t͡ʃ]. Rare in Kochetov's survey are Type II and III processes targeting labials, which occur in at most one language family each, as well as Type III processes targeting dorsals, which occur in only two language families. Like Bateman, Kochetov finds that place-changing (Type II or III) palatalization of labials implies place-changing palatalization in both coronals and dorsals. While palatalization processes targeting only coronals or only dorsals are common, processes targeting only labials are unattested.

5.3.1 Palatalization of labials in Setswana

Setswana is rare among the world's languages in that it exhibits a synchronic alternation between labials and palatoalveolars (full palatalization), sometimes to the exclusion of coronal and dorsal segments. The existence of full palatalization of labials in Setswana, as well as the opacity of its triggers, has sparked much debate regarding the diachronic origins of this alternation in particular, as well as the nature of palatalizing processes more

generally. Palatalization of labials in Setswana occurs in three morphological contexts.[3] Given the tendency for palatal glides and front vowels to trigger palatalization, the most straightforward case of palatalization in Setswana is in the causative. Here, palatalization is triggered by the suffix *-ja*, as shown in (21):

(21) Palatalization of labials in the Setswana causative (Cole 1955, 45)

a. /-tɬʰalɪ**ɸ**-ja/ → [-tɬʰalɪ**tʃʰʷ**a] 'become wise'

b. /-natɪ**ɸ**-ja/ → [-natɪ**tʃʰʷ**a] 'become pleasant'

Aside from the bilabial fricative /ɸ/, only three other segments palatalize in the causative: /l, n, g/. Palatalization of other segments is avoided by selecting an alternate causative suffix, *-isa*, which is preferred in modern Setswana (Cole 1955).

In the diminutive, labials, coronals, and velars all palatalize equally before the suffix *-ana*, but the palatalizing trigger in this context is debated, given that there is no surface palatal or front vowel in the diminutive suffix. Palatalization of labials in this environment is shown in (22):

(22) Palatalization of labials in the Setswana diminutive (Cole 1955, 43-44):

a. /tɬʰa**p**i-ana/ → [tɬʰa**tʃ**wana] 'a small fish'

b. /tsʰɛ**pʰ**ɛ-ana/ → [tsʰɛ**tʃʰ**wana] 'a small springbok'

c. /mʊχʊ**b**ɪ-ana/ → [mʊχʊ**dʒ**wana] 'a small pan or pond'

d. /mʊra**ɸ**ɪ-ana/ → [mʊra**tʃʰ**wana] 'a small nation or tribe'

Palatalization in the diminutive is argued to be triggered by the presence of a palatal glide, which is either inserted to resolve hiatus (Bateman 2007) or present in the underlying form of the suffix, *-jana*, a morpheme meaning 'child' that can be found in many Bantu languages (Ohala 1978). In either case, palatalization is triggered by a palatal glide present at some point in the derivation.

3. Cole (1955) describes a fourth morphological context for palatalization, following the class 3 prefix *le-*, but he reports this case to no longer be productive, and it is not generally discussed in the literature on palatalization in Setswana.

Most unusually, palatalization also occurs in the formation of passives. Setswana passives are formed by adding the suffix *-wa* to the verb, causing labialization of non-labial consonants, and palatalization of labial consonants. Palatalization of labials in the passive is shown in (23):

(23) Palatalization of labials in the Setswana passive (Cole 1955, 43):

a.	/lɔ**p**a-wa/	→	[lɔ**tʃ**^{w}a]	'be requested'
b.	/tɬ^{h}ɔ**p**^{**h**}a-wa/	→	[tɬ^{h}ɔ**tʃ**^{**h**w}a]	'be chosen'
c.	/rɔ**b**a-wa/	→	[rɔ**dʒ**^{w}a]	'be broken'
d.	/ala**ɸ**a-wa/	→	[ala**ʃ**^{w}a]	'be cured'

The fact that palatalization is triggered only on labial segments is unexpected, given that coarticulatory factors make labials the least likely place of articulation to undergo palatalization. Passive forms for coronal-final and velar-final stems are given in (24):

(24) Labialization in the Setswana passive (Cole 1955, 192):

a.	/bɔ**n**a-wa/	→	[bɔ**n**^{w}a]	'be seen'
b.	/ra**t**a-wa/	→	[ra**t**^{**w**}a]	'be loved'
c.	/rʊ**k**a-wa/	→	[rʊ**k**^{**w**}a]	'be sewn'
d.	/a**g**a-wa/	→	[a**g**^{**w**}a]	'be built'

As with palatalization in the diminutive, numerous explanations have been given for the source of palatalization observed in the passive. Although [w] is an unusual trigger for palatalization, Ohala (1978) argues that the passive originally had two forms: *-wa* and *-iwa*, the latter of which appeared following labials due to a ban on labialization of labials.

Although a palatalizing trigger can be posited in all three cases, the question remains as to how palatalization came to affect labial segments, particularly where coronal and dorsal segments were unaffected. Explanations fall into two camps: 1) a direct change from labial to palatal as the result of misperception, or 2) telescoping, such that palatalization arose gradually through a series of independent sound changes. Ohala (1978) argues for the former,

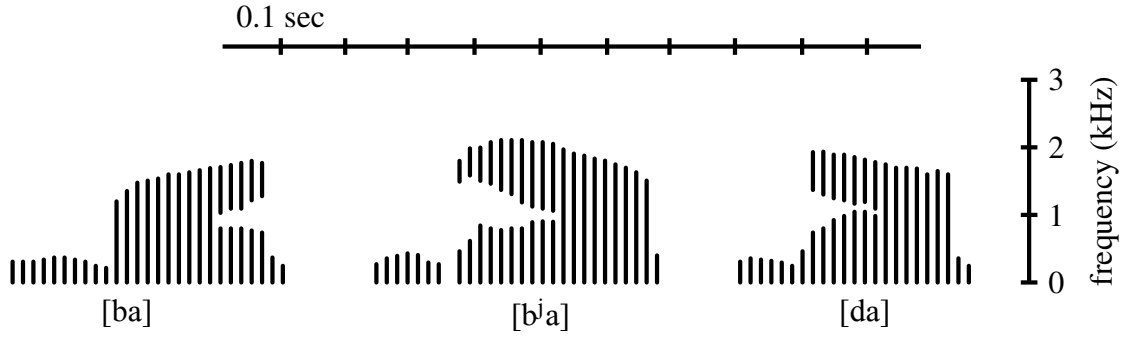
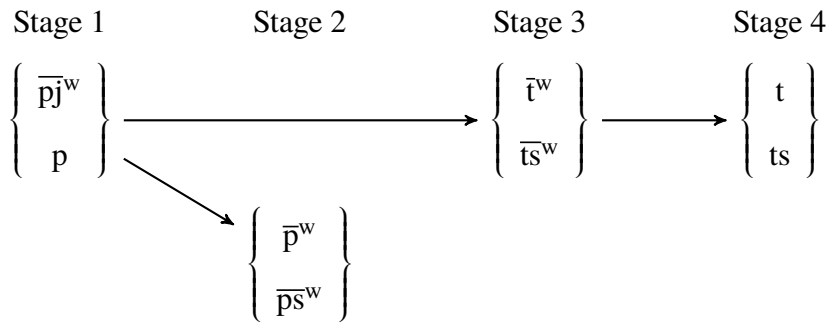


Figure 5.2: Spectrogram tracings for labial, palatalized labial, and coronal stops in Russian. From Ohala (1978, 374). Data originally from Fant (1960).

suggesting that these changes can be explained by the acoustic-perceptual properties of palatalized labials. This argument is based largely on evidence from Fant (1960), shown in Figure 5.2. Ohala notes that the F2 transition for [bʲa] is more similar to that of [da] than it is to [ba], which could cause the listener to mistake the labial stop for a coronal, particularly if the stop burst is missed. He also suggests that experimental perception data support this hypothesis, based on studies by Lyublinskaya (1966) and Winitz, Scheib, and Reeds (1972). Winitz, Scheib, and Reeds (1972), for example, find that /pi/ is reliably perceived as /pi/ when listeners are presented with the stop burst alone, but that when the stop burst is combined with 100 ms of the following vowel, listeners misperceive /pi/ as /ti/.

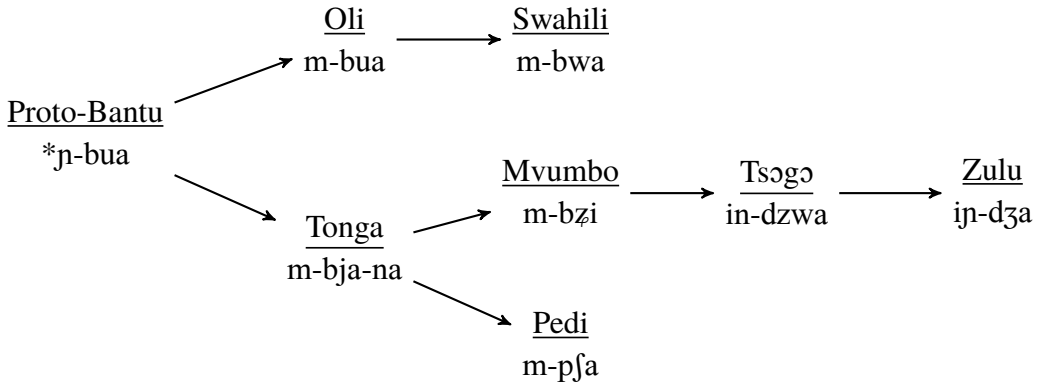
Ohala proposes the following progression for the palatalization of labials, where Stage 2 is a "possible but not necessary intermediate stage" between Stages 1 and 3:

(25) Progression of the palatalization of labials (Ohala 1978, ex. 19):

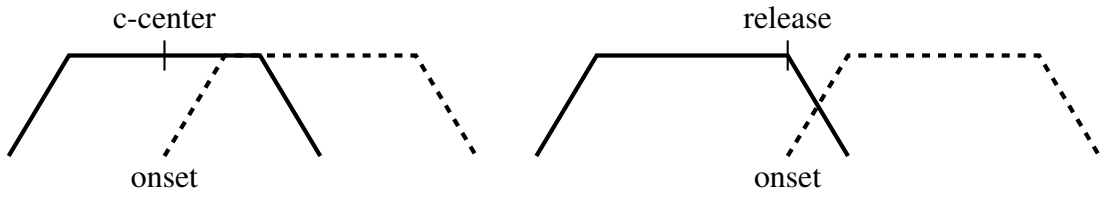


In this schema, labialization spans both the stem-final consonant and the palatalizing suffix (indicated by the bar above these segments), accounting for the retention of rounding on the palatal in words like *-tʰalɪtʃʰʷa* 'become wise', seen in (21) above. The existence of Stage 2 comes from 'old-fashioned' Setswana words like *boɸa* 'tie on back', which, when made passive, becomes *boɸʃʷa*. While this might be viewed as evidence for a telescoping account, Ohala argues that this stage is not necessary, based on the perceptual results described above. In later stages, labialization may be lost entirely, such that *-tʰalɪtʃʰʷa* is realized as *-tʰalɪtʃʰa* (Cole 1955). Ohala finds support for this progression from a survey of Bantu more generally:

(26) Reflexes of palatalized labials across the Bantu family (Ohala 1978, ex. 20):



Proto-Bantu *bu* sequences remain labial in Oli and Swahili, but are palatalized to various degrees in Southern Bantu languages. Secondary palatalization is found in Tonga, while the most extreme case of palatalization with a loss of rounding is found in Zulu.



(a) Full palatalization **(b)** Secondary palatalization

Figure 5.3: Schematic representation of full and secondary palatalization. Solid line indicates consonantal gesture and dashed line indicates vocalic gesture. Adapted from Bateman (2007).

Bateman (2007, 2010), however, provides a different interpretation of the Bantu data, arguing that the change did not occur in a single step, but is instead a case of historical telescoping (Hyman 1975; Kenstowicz and Kisseberth 1977). Bateman argues that palatalization of labials is not synchronically motivated, providing an Optimality theoretic account (Prince and Smolensky 1993) grounded in Articulatory Phonology (Browman and Goldstein 1989). She argues that the rarity of full palatalization of labials, as well as the differences between full and secondary palatalization, can be understood through the mechanisms of gestural coordination and gestural overlap. On this account, the separate processes of full and secondary palatalization arise as a result of differences in the timing of the consonantal and vocalic gestures. In the case of full palatalization, the onset of the vocoid is aligned with the the c-center of the preceding consonantal gesture, but in the case of secondary palatalization, the onset of the vocoid is aligned with the release of the consonantal gesture. This is schematized in Figure 5.3.

While the gestural timing is the same for all segments regardless of their place of articulation, differences between labial and non-labial segments arise due to differences in the articulators involved. In the case of coronal and dorsal consonants, both the consonantal and

	Passive forms			Diminutive forms		
-p	tʃ^{w}	pʃ^{w}	ps^{w}	tʃ^{w}	pʃ^{w}	pʃ
-b	dʒ^{w}	bdʒ^{w}		dʒ^{w}, dʒ	bdʒ^{w}	bdʒ
-p^{h}	tʃ^{hw}	pʃ^{hw}		tʃ^{hw}	pʃ^{hw}	pʃ^{h}
-ɸ	ʃ^{w}	ɸʃ^{w}		ʃ^{w}		

Table 5.2: Dialectal forms of palatalized labials in Setswana. From Cole (1955), cited by Bateman (2007).

vocalic gestures are lingual. When these gestures overlap, a single, blended gesture is the result. On the other hand, palatalization of labials involves gestures of two separate articulators: the labial consonantal gesture, and the lingual vocalic gesture. When these gestures overlap, the result is a partial blocking of the lingual gesture by the labial gesture. A blended gesture is impossible, because as independent articulators, movements of the lips and tongue do not perturb one another. Although it is possible for the onset of the vocoid to align with the c-center of the labial gesture, this overlap does not result in a palatal consonant. Instead, an overlap of the gestures for /p/ and /j/ results in a sound like [pi], rather than [t͡ʃ] or [p^{j}].

In addition, where historical or dialectal data is available, there is evidence that full labialization does not typically occur as a direct sound change. This can be observed through intermediate forms found not only in Bantu, as seen in (26), but also in dialects of Polish (Kochetov 1998). For instance, Cole (1955) provides a list of dialectal forms of palatalized labials in Setswana, given in Table 5.2. These data show that palatalization is not the only outcome of labial + palatal sequences in Setswana. Rather, 'palatalized' labials can also arise as sequences of a labial stop plus a palatoalveolar fricative, or as a rounded affricate. Similar data are reported by Guthrie (1970), who shows that numerous reflexes of Proto-

Bantu **pi* and **pu* sequences consist of a labial stop plus a palatal fricative or glide. Some examples are given in (27):

(27) Intermediate stages in the palatalization of Proto-Bantu **p* (Guthrie 1970):

	Proto-Bantu		Modern Form	
Tumbuka	*-**pí**à-	→	-**pɕ**a	‘new’
	*-**pí**ágid-	→	-**pɕ**eɽ-	‘sweep’
Tonga	*-**pí**à-	→	-**phj**a	‘new’
Yao	*-**pí**ágid-	→	-**pj**āɟil-	‘sweep’
Maŋanja	*-**pí**yò	→	im|**pɕ**o	‘kidney’
	*-**pí**ù-	→	-**pɕ**u	‘red’
S. Sotho	*-**pù**ànj-	→	-**pɕ**hatl’	‘pound’
Pedi	*-**pu**any-	→	-**pɕ**anj	‘pound’

Kochetov (1998) provides examples from several dialects of Polish, which are similar to those seen in Bantu. These data are presented in (28):

(28) Plain-palatal alternations in Polish dialects (Kochetov 1998):

	Plain	Palatalized				
Dialects:	All	Bartki	Kręgi Stare	Mrągowo	Jabłonka	
a.	ła[p]a	ła[pʲ]e	ła[pj]e	ła[pç]e	ła[pɕ]e	‘paw’
b.	ro[b]ota	ro[bʲ]i	ro[bj]i	ro[bʝ]i	ro[bʑ]i	‘work’/‘make’

In all dialects, plain stops appear before non-front vowels and word-finally. However, the realization of palatalized labials, which occur before front vowels, varies across dialects. A palatalized labial can be produced with secondary palatalization, as in Bartki; with a full palatal glide, as in Kręgi Stare; with a nonstrident palatal fricative, as in Mrągowo; or with a strident prepalatal fricative, as in Jabłonka. These variants appear not only as alternants before front vowels, but also in non-derived environments, as observed in (29):

(29) Non-derived palatalized labials in Polish dialects (Kochetov 1998):

Dialects:	Bartki	Kręgi Stare	Mrągowo	Jabłonka	
a.	[pʲ]ivo	[pj]ivo	[pç]ivo	[pɕ]ivo	‘beer’
	[pʲ]otr	[pj]otr	[pç]otr	[pɕ]otr	‘Peter’
b.	[bʲ]ały	[bj]ały	[bʝ]ały	[bʑ]ały	‘white’
	ko[bʲ]eta	ko[bj]eta	ko[bʝ]eta	ko[bʑ]eta	‘woman’

If these data are interpreted as representing different stages of a sound change in progress, or as sound changes which were ‘completed’ to varying degrees, a telescoping account of palatalization of labials in Setswana is generally supported. While none of the Polish dialects given here exhibit full palatalization of labials, each exhibits some degree of glide strengthening, with the labial retained. In order to achieve full palatalization of labials, then, all that is required is for the labial to independently delete, which might occur as the result of lenition. Indeed, Bateman argues that palatalization of labials in modern Setswana is in fact a reflection of three historical changes: secondary palatalization of labials before a front vowel or glide, hardening of the palatal glide, and deletion of the labial. These stages are presented in (30):

(30) Stages of palatalization of labials in the causative (Bateman 2007):

-lap- + ja → -lapja → -lapʃa → -laptʃʷa → [-latʃʷa] ‘make sb. tired’

-leoɸ- + ja → -leoɸja → -leoɸʃa → -leɸtʃʷa → [-leotʃʷa] ‘make sb. sin’

While this progression is supported by dialectal and historical evidence, however, Bateman does not provide much discussion of why these changes should occur in the first place (i.e., why the post-labial glide should harden, and why the labial should delete), nor does she provide experimental data to refute Ohala’s (1978) misperception-based account.

Kochetov (2016) provides a synchronic account of palatalization in Kirundi, arguing that glide hardening and palatalization are competing repair strategies for marked C+pal sequences. Importantly for the question at hand, he also provides acoustic data from Korean

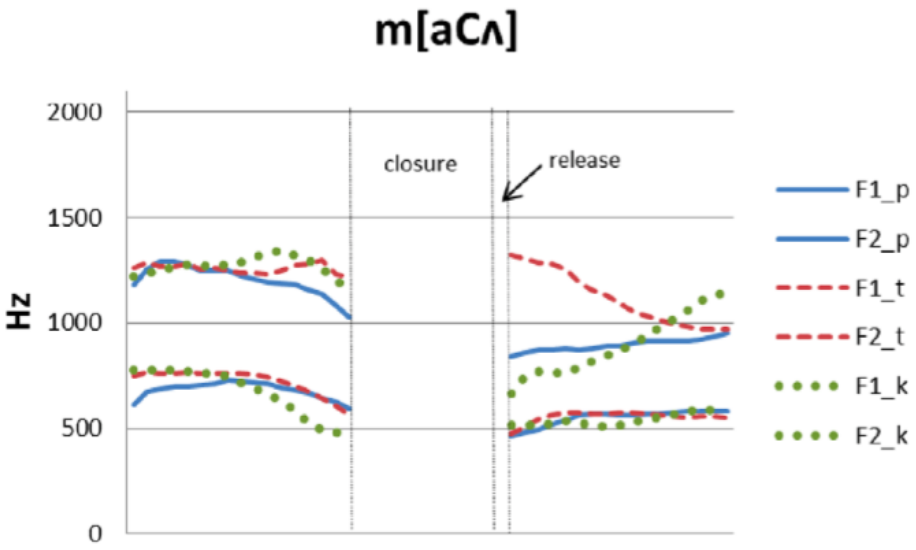


Figure 5.4: Mean F1 and F2 of plain stops in Korean. From Kochetov (2016), licensed under CC BY 4.0.

C+j sequences on the acoustic correlates of palatalization. Figure 5.4 shows the mean F1 and F2 values for vowels preceding and following plain (non-palatalized) stops in Korean. It is observed that the primary difference between labial, coronal, and velar stops is the frequency of F2 following the stop closure. While the F2 transition out of the closure rises for labial and velar stops, F2 falls following a coronal stop. This is unsurprising given the association of anterior lingual constrictions with a high F2, and of labial and dorsal constrictions with a low F2, as captured by the feature [grave] (Jakobson, Fant, and Halle 1951) and the Peripheral node (Rice 1995). In contrast, F1 and F2 transitions for palatalized stops are given in Figure 5.5. It can be seen that before a palatal glide, labial and velar stops no longer lower the value of F2; instead, F2 begins high for all places of articulation, and falls throughout the transition into the vowel. Even though there are slight acoustic differences between the three onsets in the transition from the preceding vowel, it is clear that the

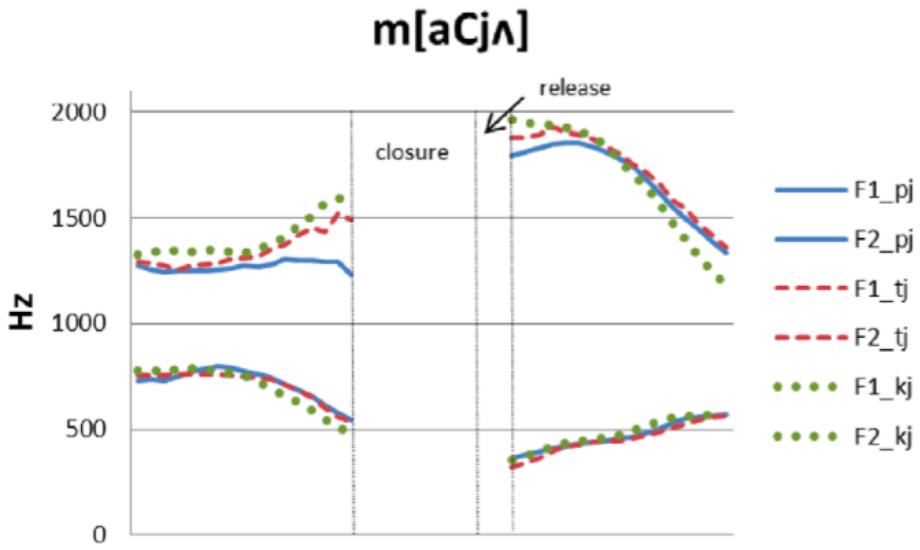


Figure 5.5: Mean F1 and F2 of stop + palatal sequences in Korean. From Kochetov (2016), licensed under CC BY 4.0.

acoustic similarity of palatalized labial, coronal, and velar stops would present a perceptual challenge for listeners on a purely auditory basis.

Given the existence of intermediate forms within dialects of Setswana, across Bantu, and in Polish, the telescoping account of full palatalization of labials in Setswana argued for by Bateman (2007, 2010) is plausible. Yet, the question of acoustic similarity and misperception remains. If palatalized labials are acoustically similar to palatals, why aren't they misperceived as such? Bateman (2007, 2010) provides three arguments against Ohala's (1978) misperception-based account. First, she argues that acoustic similarity does not imply perceptual similarity. While this may be true in principal, acoustic similarity is widely appealed to in the literature as a source of sound change, and has proven sufficient to explain many common sound changes. The fact that sound changes frequently have origins in acoustic similarity has led Ohala (1989) to state that "what looks similar to the eye [in spectrograms]

will sound similar to the ear and thus be subject to confusion" (183). In order to falsify a misperception-based account of palatalization of labials, it must either be shown that secondarily palatalized labials are in fact *not* acoustically similar to palatals, or that some other factor prevents them from being misperceived. Given acoustic data from Fant (1960) and Kochetov (2016), the former argument seems untenable. This suggests that other factors, visual speech cues being one possibility, inhibit misperception of labials with secondary palatalization.

Second, Bateman (2010) notes that a misperception-based approach predicts bidirectional change. That is, if palatalized labials can be misperceived as palatals, it should also be the case that palatals are misperceived as palatalized labials. Although asymmetries in misperception-based change are not entirely well understood (Garrett and Johnson 2011), the literature on sound change provides many examples of asymmetric changes, such as full velar palatalization, in which /ki/ sequences become [t͡ʃi] (Guion 1998), or diachronic *th*-fronting, where /θ/ becomes [f] (McGuire and Babel 2012). One explanation for this type of asymmetry comes from Chang, Plauché, and Ohala (2001), where it is argued that listeners are more likely to miss an existing cue than to mistakenly perceive one that does not exist. Although [ki] and [ti] have similar formant structures, [ki] exhibits an additional mid-frequency peak that [ti] does not. They argue that listeners are more likely to miss this cue in [ki] than to hear it in [ti], so /ki/ > /ti/ is a more common sound change than /ti/ > /ki/. They compare this asymmetry to a study of the misperception of capital Roman letters, where 'Q' was more often misperceived as 'O' and 'E' as 'F' than the reverse (Gilmore et al. 1979).

Finally, Bateman (2010) argues that misperception-based approaches can not account for the cross-linguistic rarity of full palatalization of labials. Numerous researchers have argued that typological patterns can be predicted by universal properties of speech production and perception, including Ohala (1983, 1989, 1993) and Blevins (2004), among many others. If palatalized labials are easily misperceived as palatals, it is expected that full palataliza-

tion of labials will be a common sound change. This is especially true given that labials with secondary palatalization are the most frequent type of palatalized consonant (despite occurring only when the language also has secondary palatalization of coronals or dorsals). With such a large number of languages exhibiting secondary palatalization on labials, and given the acoustic similarities between palatals and labials with secondary palatalization, one would expect that misperception should occur in at least some of these cases, giving rise to a greater number of synchronic alternations between labial and palatal segments than is actually observed. However, the perceptual integration of visual speech cues, as shown for vowels in Chapter 4, may provide an explanation for the typological rarity of full palatalization of labials. Despite the fact that labials with secondary palatalization and full palatals exhibit a high degree of acoustic similarity, they differ substantially in terms of their visual correlates. Because labials with secondary palatalization provide a clear visual cue to the listener that their place of articulation is labial, language learners can easily recover the sound's place of articulation, thereby avoiding misperception.

5.4 Chapter Summary

The three types of sound change discussed in this chapter have provided support for the hypothesis that in the course of diachronic sound change, labial segments tend to remain labial. Such a tendency was argued for by Ohala and Lorentz (1977), but this generalization concerned only labiovelar segments, and was not extended to other types of sound change. In fact, Ohala (1978) argued elsewhere that palatalization of labials was both common and easily understood under a misperception-based account of sound change. Recent work by Bateman (2010) and Kochetov (2011), however, shows that palatalization of labials is quite rare, in contrast to palatalization of coronal and dorsal segments, which is frequent. In addition, work on debuccalization shows that the debuccalization of labial segments is rare,

except when it occurs in the context of a neighboring round vowel, or when an existing pattern of coronal or dorsal debuccalization is extended to labials.

As noted above, labial consonants exhibit several properties that explain why they behave differently from coronal and dorsal consonant in these three patterns of sound change. Because the lips are located at the forwardmost point of the vocal tract, they can block more posterior articulations; because the lips are independent from other articulators, they are both immune to certain types of coarticulation and can exhibit an extended temporal domain; and finally, labial articulations are visually salient. The patterns observed for debuccalization of labials and labial-velar alternations can potentially be explained by any of these properties, and the same holds for the auditory and visual affinity of velar nasals and nasalized vowels investigated by Johnson, DiCanio, and MacKenzie (2007). For these changes, visual cues may influence the direction of change, but these patterns can also be accounted for on the basis of acoustics and auditory perception alone. However, in the case of palatalization of labials, the auditory and visual inputs provide conflicting information. While the acoustic similarity of palatals and labials with secondary palatalization suggest that labials with secondary palatalization should be susceptible to misperception, as is the case for coronals and dorsals with secondary palatalization, the visual input provides a clear cue to their place of articulation. The results of the perception experiment in Chapter 4, as well as findings on audiovisual speech perception from the literature more generally, suggest that the incorporation of audiovisual speech cues by listeners (and language learners) may help to resolve ambiguities in the speech signal, thereby inhibiting some types of misperception-based sound change.

Chapter 6

Discussion and Conclusion

The experiments presented in this dissertation have provided a closer look at the articulatory and perceptual factors that constrain patterns of articulatory variation and diachronic sound change. In Chapter 2, it was shown that speakers of two varieties of American English achieve back vowel fronting by tongue fronting rather than by unrounding the lips, even though both articulatory strategies would result in an increase in F2. For one group of speakers, from Southern California, this result may be attributed to the relationship of back vowel fronting to the coarticulatory effects of preceding coronal consonants. Even for speakers from South Carolina, however, where the fronting of /u/ and /o/ is advanced after non-coronal onsets, the back vowels retain their rounding as they undergo fronting.

Chapter 3 considered a case of unconditioned back vowel fronting, the fronting of /ɑ/ and /ɔ/ that is characteristic of the Northern Cites Shift. Historically, speakers in the Inland North have retained the contrast between /ɑ/ and /ɔ/, given that the fronting of /ɑ/ preceded the fronting of /ɔ/. However, the strength of the acoustic contrast between these two vowels varies by speaker age, partially as a result of the backing of /ɑ/ among younger speakers. Articulatory data show that the speakers with the strongest COT-CAUGHT contrast distinguish the vowels with differences in both tongue position and lip rounding. For speakers with a relatively weak COT-CAUGHT contrast, these vowels are typically distinguished by lip rounding alone. In the careful speech task, it was shown that the majority of speakers enhance the lip rounding distinction between /ɑ/ and /ɔ/, providing support for the hypothesis that speech is optimized for visual perceptibility. However, many speakers also enhanced the acoustic

difference between these vowels, as the well as the distinction between the two vowels in terms of tongue shape. It was argued that articulatory configurations in which /ɑ/ and /ɔ/ are contrasted solely through a difference in tongue position (with no lip rounding distinction) are dispreferred due to the absence of visual lip rounding cues.

Chapter 4 explicitly tested this hypothesis through a perception experiment investigating audiovisual identification of /ɑ/ and /ɔ/. Tokens containing auditory /ɑ/ and /ɔ/ were presented to participants, paired with video containing either visually round lips or visually unround lips. It was found that perceivers who produce the COT-CAUGHT contrast with a difference in lip rounding were significantly more likely to perceive /ɔ/ as /ɑ/ when it was presented with video of unround lips. This result supports the hypothesis that unround variants of /ɔ/ are perceptually weaker than round variants. While unround variants may provide acoustic contrast with /ɑ/, visually round variants provide contrast in both the auditory and visual domains. Thus it is predicted that speakers will prefer round variants of /ɔ/ in order to optimize the perceptibility of their speech.

Finally, Chapter 5 presented a review of sound changes involving labial segments, considering whether visual perceptibility influences the direction of diachronic sound change. Three types of sound change were examined, including labial-velar alternations, debuccalization, and palatalization of labials. Typological data show that in all three types of change, labial segments tend to remain labial, except when the labial gesture can be attributed to a neighboring labial segment. In the cases of labial-velar alternation and debuccalization, this tendency can be explained in part by both auditory and visual factors: in addition to being visually salient, labial gestures have the capacity to block more posterior articulations, as well as to extend temporally over multiple segments. In the case of palatalization of labials, however, labials with secondary palatalization have been found not to become palatal in spite of their acoustic similarity to palatals. It was argued that visual speech cues can account for this cross-linguistic tendency because they offer language learners with a

clear cue to the segment's place of articulation. Historical data thus support the hypothesis that visual speech cues can aid perceivers in distinguishing between acoustically ambiguous sounds, and that this additional perceptual modality can inhibit misperception-based sound change.

Taken together, these results suggest that visual speech cues contribute to the organization of phonological systems in at least two respects. In terms of synchronic variation, audiovisual speech cues can determine which articulatory strategy a speaker selects, when multiple articulatory configurations would result in the same acoustic output. Furthermore, audiovisual cues enhance otherwise weak acoustic contrasts, thereby inhibiting misperception-based change. As a result, labial segments often retain their place of articulation during diachronic change. The theoretical implications of these findings are addressed in the next section, followed by a discussion of some outstanding questions and directions for future work. The final section of this chapter concludes.

6.1 Theoretical Implications of Audiovisual Speech Perception

The question raised in the abstract of this dissertation asks, what are the factors that shape linguistic sound systems? As discussed in Chapter 1, a number of competing proposals have been put forth. Some researchers, such as Ohala (1981, 1993) and Blevins (2004) have focused on the transmission of language from one generation of speakers to the next. In these theories, the listener plays a central role in interpreting a speaker's intended message; when this interpretation fails, sound change can arise, but such changes are not intended by either the speaker or the listener. Others have focused on the role of the speaker in attempting to make themselves understood by the listener, even under adverse communicative conditions. For instance, Lindblom (1990) and Hayes, Kirchner, and Steriade (2004) argue that speakers adapt their production targets based on both internal, speaker-oriented goals, as

well as external, listener-oriented goals. On the one hand, speakers have a natural tendency to produce sounds with as little articulatory effort as possible. On the other hand, given that speakers also wish to be understood by their interlocutor, they have a competing impetus to speak clearly enough for their message to be successfully conveyed. This type of theory makes predictions both for online speech production, as well as language acquisition (and thereby language change). In terms of production, speakers are predicted to optimize their speech targets for maximum perceptibility, particularly when communicating under noisy conditions. In terms of acquisition, phonetically marked or unnatural sound patterns are predicted to be more difficult to learn, and are therefore more likely to undergo change due to reanalysis by the learner. Both aspects suggest that natural or unmarked patterns will be preferred over unnatural or marked patterns in the course of language change. Diehl and Kluender (1989) argue against gesture-based theories of speech perception, and propose that phonological systems are organized around optimal auditory-acoustic properties, such as maximal distance in the vowel space. One aspect central to all of these approaches is the perceptibility of the speech signal, and the extent to which listeners are able to recover the speaker's intended message. Each of these approaches, however, considers perception only insofar as it relates to auditory perception, despite the fact that the integration of non-auditory modalities is well known in the speech perception literature. Although some theories of speech perception itself, such as the gestural approaches of Motor Theory (Liberman and Mattingly 1985) and Direct Realism (Fowler 1986) take visual speech perception to be an important source of evidence, these theories are not typically considered with respect to their implications for language variation and change.

The preceding chapters have shown that there are several ways in which visual speech cues can influence the organization of phonological systems; two questions to consider with respect to the theories just described are as follows. First, what are the ways in which speech is optimized? Is it optimized for auditory perceptibility, visual perceptibility, or both?

Second, at what stage does this optimization occur? Does optimization arise because sounds that are easier to perceive are simply more likely to be correctly interpreted by the listener? Or, do speakers actively optimize their speech for perceptibility?

Regarding the first question, Diehl and Kluender (1989) argue explicitly against the notion that vowel systems are dispersed in the articulatory domain. As noted in Chapter 1, they observe that vowel systems composed of the vowels /i u a/ are cross-linguistically common, while vowel systems composed only of /y ɯ a/ are unattested. They account for this fact by explaining that while the vowel pairs /i-u/ and /y-ɯ/ are equally distinct in terms of articulation, the complementary acoustic effects of backing and rounding mean that only the pair /i-u/ is maximally distinct in terms of acoustics. A vowel that is [+back, +round] offers a better acoustic contrast with one that is [−back, −round] than a [+back, −round] vowel does with a [−back, +round] vowel. This observation is certainly correct, and a large body of research has shown that acoustic distance (and thus auditory distinctiveness) plays an important role in the organization of vowel systems (Liljencrants and Lindblom 1972; Lindblom 1986; de Boer 2001; Flemming 2004).

However, the account proposed by Diehl and Kluender (1989) offers no prediction as to how vowels should behave when they undergo change due to external pressures. In the case of non-low back vowel fronting analyzed in Chapter 2, the fronting of /u/ is caused not by pressure for this vowel to become more distinct from any other, but, in fact, the opposite: the coarticulatory influence of a preceding coronal interferes with the acoustic realization of /u/, causing it to become *less* distinct from /i/. When acquiring a vowel system with fronted /u/, the language learner must then choose some articulatory strategy for producing this acoustically weak vowel, whether it is realized as [y], [ɨ], [ʉ̟], or [ɯ̟]. The results of the production experiment show that the vowel speakers tend to produce is more like [y] or [ʉ̟], which are round, rather than [ɨ] or [ɯ̟], which are unround. In the case of the Northern Cities Shift investigated in Chapter 3, the fronting of /ɔ/ apparently *is* driven by the demands of

acoustic contrast, as it occurs as part of a chain shift whereby /ɔ/ undergoes fronting in order to fill an open region of the vowel space caused by the fronting of /ɑ/. However, this pressure for acoustic dispersion cannot predict the articulatory strategy used to front /ɔ/, because both tongue fronting and lip unrounding have the same effect on the acoustic signal. Either strategy can therefore be used to fill the gap in the acoustic vowel space. Again, the results from the articulatory study of the COT-CAUGHT contrast show that speakers generally retain the lip rounding gesture for /ɔ/. The results from the two production experiments presented in this dissertation thus show that, in both cases, speakers tend to preserve lip rounding as vowels undergo fronting. While a speaker could just as easily reach their acoustic target by unrounding, such a strategy would not be motivated by visual perceptibility. By achieving an increase in F2 through fronting the tongue, however, the visual distinctiveness of the vowel is preserved, even though the auditory distinctiveness has been weakened by external pressures. The results of Chapter 4 show that this visual distinctiveness can play a critical role in the perception of vowel contrast.

While these results suggest that sound systems are in fact optimized for visual perceptibility in some respects, it is not clear whether this optimization is intentional, as predicted by teleological models like H&H theory. Ohala's model of innocent misapprehension and Evolutionary Phonology would instead posit that speakers retain lip rounding for vowels that undergo fronting simply because a) speakers are imitating the articulatory patterns of their predecessors, or b) round variants are easier to perceive and only easily perceived sounds are likely be passed on. It would follow from the latter position that unround variants could arise by chance, but that the contrast would subsequently collapse if the loss of visible lip rounding caused the vowel to become perceptually weak. The hypothesis that speakers actively optimize their speech for visual perceptibility was tested in Chapter 3. In that experiment, participants were asked to repeat a series of sentences designed to elicit careful speech, in which words containing /ɔ/ were produced alongside a minimally con-

trastive word containing /ɑ/ and vice versa. The acoustic results showed that nearly all (13 of 15) of the participants increased the degree of acoustic contrast between /ɑ/ and /ɔ/, as indicated by an increase in the Pillai score measure of acoustic distance. The increase in Pillai score was quite dramatic for many of these participants, particularly for those who produced these vowels with only a moderate degree of contrast in the normal speech task. The articulatory results showed that the majority of speakers increased the difference between /ɑ/ and /ɔ/ in terms of lip rounding/spread, suggesting that speakers enhance their speech for visual perceptibility, at least to some extent. Overall, ten of the fifteen participants increased the lip rounding difference between /ɑ/ and /ɔ/ when producing careful speech. Some of these speakers did so by significantly decreasing the degree of lip spread for /ɔ/, while others increased the degree of lip spread for /ɑ/. Two speakers modified the degree of lip spread for both vowels. Five speakers failed to increase the lip rounding distinction, but for four of these speakers, this finding was explained by insufficient measurement techniques or by idiosyncratic factors.

It was argued that this finding provides support for the hypothesis that speakers optimize their speech for visual perceptibility. However, these results must be interpreted with some caution; for many of these speakers, an increase in the lip rounding distinction between /ɑ/ and /ɔ/ might simply be attributable to increasing the acoustic distance between these vowels. This could be said, for instance, of CHI013, the sole speaker who produced a lip rounding contrast only in the contrastive speech task, but not in the normal speech task. Because this speaker produced these vowels with distinct tongue shapes in the normal speech task, she may have recruited lip rounding in the contrastive speech task because it was the least effortful way to increase the acoustic contrast. This speaker was observed to show the greatest increase in Pillai score among all the speakers in the study, an increase of 0.318. Her performance in the perception experiment also supports the notion that she was optimizing her speech for auditory, not visual, perceptibility, because she was herself insen-

sitive to a loss of visible lip rounding cues for /ɔ/. For speaker CHI012, who also showed a substantial increase in Pillai score, the increase in lip rounding for /ɔ/ was also accompanied by a tongue position difference not observed in normal speech; both tongue and lip contrasts may have been necessary to achieve such a large increase in acoustic distance. It cannot strictly be said that these speakers are optimizing their articulations for visual perceptibility, since increased lip rounding may just be a requirement for maximizing auditory perceptibility.

Several of the speakers in this study, however, do provide some evidence that articulatory patterns can be optimized for visual perceptibility alone, with only a small or no increase in acoustic distance. CHI003 was one of two speakers in the study who produced a smaller degree of acoustic contrast between /ɑ/ and /ɔ/ in careful speech, yet who nevertheless showed a significant increase in the degree of lip rounding (i.e., a decrease in lip spread) for /ɔ/. Another speaker, CHI018, also showed a significant increase in the degree of lip rounding for /ɔ/, but an increase in Pillai score of 0.08, which was one of the smallest Pillai score increases observed in this study. Interestingly, this speaker produced a tongue position distinction in the normal speech task, but lost this distinction in careful speech. There is no clear acoustic motivation for eliminating the tongue position distinction; indeed, the loss of the tongue distinction appears to counteract much of the acoustic enhancement afforded by the increase in lip rounding. This pattern can be understood under the framework of H&H theory, however, as one solution to the conflict between articulatory ease and maximal perceptibility. If a speaker believes that the listener is relying on visual lip rounding cues, they can increase perceptual distance between /ɑ/ and /ɔ/ by enhancing lip rounding for /ɔ/, while simultaneously reducing articulatory effort by avoiding the pharyngealized tongue position that a low back vowel typically requires. The loss of a distinction in tongue position means there is only a small net gain in acoustic contrast, but the enhanced visual cues provided by

increased lip rounding mean that the overall perceptual distance is much greater in careful speech than in normal speech.

A similar pattern is observed for speaker CHI002. This speaker was classified as producing no increase in the lip rounding contrast, but it was suggested that this may have been a shortcoming of the classification system used; despite the lip spread measurements for both vowels decreasing in parallel, the difference in mean lip spread between /ɑ/ and /ɔ/ for this speaker shows an overall increase. In the normal speech task, the mean lip spread measurement was 153 pixels for /ɑ/ and 92.8 pixels for /ɔ/, a difference of 60.2 pixels. In the careful speech task, on the other hand, the mean lip spread for /ɑ/ was 144 pixels and the mean lip spread for /ɔ/ was 67.4 pixels, a difference of 76.6 pixels. This simplified metric does not take variance around the mean into account, but nevertheless shows that, on average, /ɑ/ and /ɔ/ are more different in terms of lip spread in careful speech than in normal speech. However, while this speaker does appear to enhance the visual lip rounding distinction in her careful speech, she produces an even smaller increase in Pillai score (0.03) than CHI018 and likewise loses a tongue position distinction that was observed in normal speech. Again, this result demonstrates the trade-off between the speaker-oriented and listener-oriented demands placed on the speaker. One way to resolve the conflict between these demands is to focus articulatory enhancement efforts on the lips, because doing so will provide the greatest overall increase in audiovisual perceptibility. Articulatory effort can then be conserved by eliminating an articulatory distinction (such as tongue position) that contributes less to the overall audiovisual signal. In sum, although it is difficult to disentangle visually-oriented enhancement from acoustically-oriented enhancement, the behavior of these three speakers suggest that in some cases, speech can be optimized for visual perceptibility alone. There appears to be a fairly wide range of individual differences in enhancement strategy, however, and the extent to which speakers optimize their speech for visual perceptibility

may have been underestimated due to shortcomings of the experiment that are addressed in the next section.

Returning to the question of the speaker's role in phonological optimization, the findings of this dissertation suggest that while misperception of ambiguous speech sounds is possible, speakers also attempt to enhance their articulatory patterns for audiovisual perceptibility in order to avoid misperception. One theoretical approach that is particularly appropriate for these findings is that of Lindblom et al. (1995). Lindblom and colleagues propose a hybrid model of sound change, incorporating aspects of both the optimizing and non-optimizing approaches to sound change. They acknowledge that sound changes can occur as a result of accidental misperception of the speech signal, but also argue, following H&H theory, that the speaker plays an important role in evaluating and selecting phonetic variants for production. Central to their proposal is the speaker's knowledge of the range of possible phonetic variation in their language. Given that language users possess a great deal of knowledge about the structure of their language, Lindblom and colleagues note that many aspects of linguistic communication are predictable, even if the quality of the speech signal is degraded due to noise or other factors. This predictability allows listeners to rely on their expectations of the message in order to overcome many of the challenges inherent in speech perception. As an example, one type of information listeners can use to disambiguate the speech signal is syntactic structure. In the case of the COT-CAUGHT contrast, the nouns *bot* and *cot* are unlikely to be misperceived as the verbs *bought* or *caught*, because they appear in distinct syntactic positions. As such, if a speaker produces a perceptually non-optimal unround variant of the vowel /ɔ/ in *caught* (perhaps due to lenition, coarticulatory pressure, or production error), the listener is nevertheless likely to correctly perceive the speaker's message and will map this novel variant to their phonetic representation for /ɔ/ rather than to /ɑ/. Having incorporated this variant into their representation for /ɔ/, the listener-turned-

speaker can evaluate it for articulatory and perceptual fitness, and decide whether and when to adopt it in their own speech.

A variety of other factors, including lexical frequency (Hooper 1976; Bybee 2000) and sociophonetic knowledge (Sumner and Samuel 2009) can also influence listeners' expectations and help to resolve ambiguities in the speech signal. Because these types of knowledge require experience with the language, however, early-stage learners may be less successful in uncovering the speaker's intended message. Krakow et al. (1988), for instance, show that American English-speaking listeners perceive nasalized vowels as lower than oral vowels, but only when they appear in an oral environment (e.g., [b_d] as opposed to [b_nd]), in which coarticulatory nasalization is not expected. They argue that this is because American English-speaking listeners have experience perceiving the height of vowels with coarticulatory nasalization, but lack experience perceiving the height of vowels with distinctive nasalization. When a naïve learner encounters a nasalized vowel in a context where the coarticulatory source of nasalization is weak or absent, they are likely to misjudge the height of the vowel. If this type of perceptual error occurs with sufficient frequency, the learner may acquire an underlying form that is distinct from that of the speaker, resulting in sound change. As acknowledged by Ohala (1993), however, the majority of perceptual errors do not result in community-wide sound change, because language learners are ultimately exposed to the speech patterns of a large number of speakers in a variety of speech contexts. Under an exemplar model of speech perception (Johnson 1997; Pierrehumbert 2001), all of these previously-experienced tokens of a given sound form part of a cloud of exemplars, which can then be drawn upon as production targets.

Where the hybrid model of Lindblom et al. (1995) improves on the non-optimizing models of Ohala (1981, 1993) and Blevins (2004), then, is in accounting for the behavior of the listener-turned-speaker after they have acquired a rich phonological representation composed of more and less perceptually and articulatorily optimal tokens. By providing a

mechanism for the speaker to weigh their choices between hypoarticulated and hyperarticulated variants, the approach proposed by Lindblom et al. (1995) explains why the majority of speakers in this study increased their use of lip rounding in producing careful speech. Although changes in both tongue position and degree of lip rounding can increase the acoustic difference between a pair of vowels that are contrastive in rounding, only changes in the degree of lip rounding can enhance both the acoustic and visual difference between the two sounds. In normal speech, both round and unround variants of /ɔ/ (for example) would be equal in terms of their acoustic output and their articulatory difficulty, but visually round variants offer the greatest perceptual salience without an increase in articulatory difficulty, leading speakers to prefer round variants. Additional experimental support is still needed to verify the hypothesis that speakers do actively optimize their speech patterns for visual perceptibility, but an optimizing model of phonology will ultimately provide the best account of speaker behavior even if the hypothesis of audiovisual optimization is disconfirmed. In focusing on the imperfect diachronic transmission of speech sounds, the approaches of Ohala (1993) and Blevins (2004) either fail to consider the role of the speaker in attempting to make themselves understood, or argue that speakers can learn any perceivable phonological pattern and play no role in the optimization of sound systems. Although these approaches can account for misperception-based historical changes of the sort discussed in Chapter 5,[1] they cannot account for the behavior of speakers in a contrastive speech task like that presented in Chapter 3. Even if the participants in Chapter 3 were found

1. Many of these changes can be considered non-optimal; it is difficult to see, for instance, how the debuccalization of a labial fricative next to a round vowel improves on the language's previous state. A hybrid model of sound change acknowledges that this type of accidental misperception or misparsing can occur, and the existence of such changes is not sufficient to reject the hypothesis that speakers optimize their own speech patterns. If a language learner fails to correctly perceive a phonetic variant, it cannot be evaluated for articulatory or perceptual fitness, so accidental, non-optimizing changes can occur when the language learner fails to perceive the full range of variation present in the language. However, under the normal scenario in which speakers do acquire a system with multiple variants, those variants will be evaluated and selected for production based on the speaker's articulatory, perceptual, and sociolinguistic goals.

to optimize their speech for acoustic contrast alone (such as through non-visible differences in tongue position), the results would still support an optimizing model of sound change because non-optimizing models only account for the transmission of language between generations and not how speakers behave in real-time communication. While both types of model are able to account for the fact that /u/ and /ɔ/ remain round as they undergo fronting,[2] only models that incorporate the speaker's desire to be understood can account for the behavior of speakers in producing careful speech. The findings from this dissertation therefore support a model of sound change like that proposed by Lindblom et al. (1995), and suggest that such a model should incorporate both auditory and visual perceptibility, and perhaps other linguistically relevant modes of perception.

6.2 Future Work

This dissertation has made progress toward considering how non-auditory perceptual cues can contribute to the organization of phonological systems, but there remain a number of unanswered questions that must be addressed in future work.

As discussed in the previous section, one major outstanding issue is the question of optimization: if speakers do in fact optimize their articulatory patterns for both auditory and visual perceptibility, it is predicted that they will increase the degree of lip rounding used in careful speech, in some cases with no associated increase in acoustic contrast. The experiment presented in Chapter 3 found that this was the case for at least some speakers, but there was a wide range of interspeaker variation in terms of contrast enhancement strategy. While some speakers did increase the magnitude of their labial gestures for /ɑ/ and /ɔ/ without increasing the acoustic contrast between these vowels, other speakers focused on enhancing

2. As noted above, a non-optimizing model of change would posit that learners might simply imitate the productions of the previous generation, or that only round variants are perceptually salient enough to be correctly perceived and acquired.

the acoustic contrast by employing a tongue position distinction not observed in normal speech. One speaker appeared not to optimize their speech for either auditory or visual perceptibility, showing no enhancement in the acoustics or in either articulatory measure. A possible explanation for these mixed results lies in the design of the contrastive speech task that was used. In that experiment, speakers were asked to repeat a set of words in carrier phrases designed to elicit contrast with a minimally contrastive word, and instructed to speak as clearly as possible. However, clear speech has a number of phonetic correlates, any of which may be recruited to enhance phonological contrasts, including increased amplitude, decreased speech rate, higher or more variable pitch, expansion of the vowel space, vowel lengthening, and a range of modifications to consonants (see, e.g., Picheny, Durlach, and Braida 1986; Johnson, Flemming, and Wright 1993; Ferguson 2004; Smiljanić and Bradlow 2005). Thus, enhancement in the degree of visible lip rounding is just one of many solutions that speakers may recruit when optimizing their speech for perceptibility. This is particularly true given that speakers were not communicating under noisy conditions, but were instead speaking in a quiet environment (a sound-attenuated booth) in which there would be no anticipated issues with auditory perception. H&H theory (Lindblom 1990; Lindblom et al. 1995) predicts that speakers will adapt their speech production efforts depending on their estimation of the listener's perceptual requirements in a given communicative context. Indeed, much of the earliest research on visual speech perception, such as that of Sumby and Pollack (1954), focused on the contributions of lip reading to speech perception under noisy conditions. Speakers in this experiment, however, had no inherent reason to suppose that their imagined interlocutor would be unable to hear their speech, and therefore no reason to eschew acoustically-oriented enhancement strategies in favor of visually-oriented strategies. If this experiment were repeated with speakers completing the same task in a noisy environment, they may be more likely to focus on enhancing visual lip rounding cues than on using differences in tongue position to enhance the acoustic contrast. A related issue

with the design of this experiment is the lack of audience. Participants in this study were not speaking in a naturalistic context with an interlocutor, but were simply repeating prompts presented to them on a computer screen. The enhancement strategies observed may therefore not be reflective of speakers' typical behavior when communicating under natural circumstances. While the results from the careful speech task in Chapter 3 are informative, additional research is needed to determine the extent to which speakers enhance their articulatory patterns for visual perceptibility.

One limitation of both production experiments presented in this dissertation is the question of intra- (as opposed to inter-) speaker variation. The optimizing model of change proposed by Lindblom et al. (1995), as well as the non-optimizing models of change proposed by Ohala (1981) and Blevins (2004), take synchronic phonetic variation to be the foundation upon which diachronic sound change is built. As such, understanding the mechanisms of sound change relies to a large extent on understanding the types and range of synchronic phonetic variation that is possible in the language. With respect to the production of the COT-CAUGHT contrast, the speakers in this study were classified based on whether or not they produce a distinction between /ɑ/ and /ɔ/ in terms of lip rounding, tongue position, or both. However, this classification was made on a holistic basis, and did not consider whether an individual speaker deploys multiple articulatory configurations for the same vowel. Mielke, Baker, and Archangeli (2016), for instance, show that some speakers of American English do not use just a single bunched or retroflex variant for /ɹ/, but alternate between the two based on speaker-specific allophonic patterns. Likewise, it may be the case that speakers from Chicago alternate between round and unround variants of /ɔ/ depending on the phonological environment. It was noted in Chapter 4 and by Havenhill and Do (2018) that visual lip rounding for /ɔ/ may be difficult to perceive when it follows a labial segment like /p/ or a labialized segment like /ʃ/. While this is true to some extent for all round vowels, the effect will be stronger for /ɔ/ due to the fact that /ɔ/ is produced with a low jaw position

that makes protruding the lips more difficult (Ladefoged and Maddieson 1996). In these environments, speakers may rely more on tongue position distinctions than on lip rounding distinctions, because lip rounding, even if present, will be difficult both to produce and to perceive. Additional investigation is therefore needed to determine the extent to which the articulatory patterns for /ɔ/ vary based on the surrounding phonological environment.

In the production experiment described in Chapter 2, the articulatory configurations of individual speakers were only briefly considered, with a focus instead on articulatory patterns at the population level. The results of that experiment show that the vowels /u/ and /o/ vary to a large extent in the degree to which the tongue is fronted, with high-F2 tokens exhibiting a more fronted tongue than low-F2 tokens. However, it remains unclear whether that is also true within the speech of an individual speaker, or whether speakers tend to stick with a consistent tongue position across all tokens. The individual SS ANOVA tongue contours presented for speakers Cal007 and Cal008 suggest that there is interspeaker variation in the frontedness of the tongue positions for /u/ and /o/. Cal008 produced /u/ with a tongue position that overlapped with /i/, while Cal007 produced /u/ with a tongue position that was more back. While both of those speakers produced /u/ and /o/ with lip rounding, it could be the case that other speakers consistently use a backed tongue position for these vowels, and increase the F2 for some tokens by unrounding the lips.

Another outstanding issue relating to the experiment presented in Chapter 2 is the question of gestural timing, which would have implications for the featural representation of fronted /u/ and /o/. In the data presented here, each articulatory measure was taken at only a single point within the vowel's duration. However, /u/ and /o/ are not monophthongal, so the measurements presented here likely did not capture the full acoustic and articulatory range of these vowels. Eckert (2008) notes, for instance, that some speakers produce /u/ as a "simple fronted vowel," [y], while others produce /u/ with a front, unround nucleus, [ɪw]. Indeed, several of the participants in this study produced diphthongized variants of

/u/, with a substantial lowering of F2 in the offglide. One interesting question to investigate is whether this F2 lowering is achieved by backing the tongue or by increasing the degree of lip rounding. If it is found that the position of the tongue remains relatively front even as F2 decreases during the offglide, this would suggest that those speakers have lost the [+back] component for /u/ altogether. /u/ may therefore best be analyzed as a central or front round vowel like [ʉ] or [y] or, if diphthongized, [iʉ] or [iy]. On the other hand, if lip rounding is present throughout the vowel's duration, and F2 lowering during the offglide is achieved by backing the tongue, this would suggest that speakers have retained both the [+back] and [+round] features, but that the gestures associated with these features have undergone a temporal reorganization. Preliminary analysis suggests that both patterns are present among the speakers in this study, and that distinct featural representations may be necessary to represent both variants. While additional investigation is needed, such a finding would have interesting consequences for our understanding of the phonetics-phonology interface, and would lead to greater understanding of cross-linguistic patterns of vowel system development, such as the historical development of French /y/ from /u/.

One major constraining factor for all of the experiments presented in this dissertation is the fact that large-scale articulatory studies of linguistic variation are still in their infancy. As a result, methods for making quantitative cross-speaker comparisons and generalizations are still somewhat rudimentary. For instance, while there exist well-established and well-tested procedures for vowel formant normalization (see, e.g., Adank, Smits, and van Hout 2004), there are no clearly established techniques for normalizing measurements of tongue position or lip rounding. Moreover, acoustic studies in sociophonetics benefit from tools like FAVE-align and FAVE-extract (Rosenfelder et al. 2015), which can automate much of the measurement process. Articulatory studies have long been a mainstay in theoretical phonetics, but studies in that field are often limited to relatively small speaker samples, in many cases with only two or three speakers for a particular study. It is therefore difficult

to know the extent to which speech sounds vary in their articulation, how much of this variation can be attributed to ostensibly non-linguistic factors like physiology, and how this sort of variation is distributed among the population, among other considerations. Recent research by Noiray, Iskarous, and Whalen (2014) has found that vowels vary as much in their articulation as they do in their acoustics, and that articulatory-acoustic mappings often do not neatly fit their canonical featural descriptions. Looking at acoustics alone (or articulation alone) is therefore likely to obscure a great deal of information about how speech sounds are produced, and how speakers represent these sounds phonologically. As articulatory study in sociophonetics becomes increasingly more popular, methods for conducting large-scale studies of articulatory variation will inevitably become more sophisticated, leading to a greater understanding of the factors that constrain patterns of articulation and the role of articulatory variation in diachronic sound change.

Finally, empirical evidence is needed to support the hypothesis that visual speech cues inhibit the misperception of labials with secondary palatalization, as proposed in Chapter 5. While the results of the audiovisual perception experiment presented in Chapter 4 provide evidence that listeners rely on visual cues in perceiving labial gestures, it is not known whether visual cues are sufficient to avoid perceptual confusion of [pʲ] and [t͡ʃ].

6.3 Conclusion

The experiments presented in this dissertation contribute to our understanding of how visual and auditory speech cues interact in shaping sound systems through variation and change. Although previous work on multimodal and gestural speech perception has shown that listeners integrate a wide variety of perceptual modalities in speech perception and may directly perceive some articulatory gestures, most work in phonetics and phonology has focused on the acoustic and auditory properties of speech. This is particularly true of Ohala-

style theories of sound change, under which it is argued that acoustically similar sounds are subject to misperception and subsequent change. Such theories provide an explanation for numerous typological patterns, because acoustically similar sounds have a greater chance of being misperceived, giving rise to frequent sound change. As shown in Chapter 5, however, there exist cases in which a direct sound change between two acoustically similar sounds does not occur, such as a direct /pʲ/ > /t͡ʃ/ change, contrary to the predictions of misperception-based theories of change. The results from this dissertation suggest that these patterns can be explained in part by the availability of visual speech cues, which allow listeners (and language learners) to readily recover the primary place of articulation, thereby inhibiting sound change. Indeed, the results of the perception experiment presented in Chapter 4 show that the integration of visual cues in speech perception can enhance otherwise weak acoustic contrasts. This result is supported by patterns of articulation in three varieties of American English, which suggest that certain articulatory strategies are dispreferred on perceptual grounds, even if those strategies would result in an appropriate acoustic output. Moreover, there is evidence that speakers optimize their speech for visual perceptibility, providing support for a model in which speech sounds are subject to change due to misperception, but in which speakers can actively enhance their speech patterns to avoid being misunderstood. Consideration of both articulatory and audiovisual perceptual factors is shown to be crucial to understanding the mechanisms of sound change, uncovering patterns that cannot be explained by auditory or acoustic study alone.

Appendix A

Wordlist for Back Vowel Fronting Production Experiment

A.1 Wordlist for Production Task

	/i/	/u/	/ɪ/	/ʊ/	/e/	/o/	/ɑ/	/ɔ/
p_#					pay	Poe	pa	paw
p_p	peep	(boop)	pip		paper	pope	pop	pauper
p_t	pete	(boot)	pit	put	(bait)	(boat)	pot	
p_k	peek	pookie	pick	(book)	(bake)	poke	(bock)	(balk)
p_l	peel	pool	pill	pull	pail	pole		Paul
b_#	bee	boo			bay	Bo		
t_#	tea	too			Tay	toe	ta	(tawny)
t_p			tip		tape		top	
t_t	teat	toot	tit		Tate	tote	tot	taught
t_d	teed	tude						
t_k	teak	(duke)	tick	took	take	toke	tock	talk
t_l	teal	tool	till		tale	toll		tall
d_p	deep	dupe	dip			dope		
d_l	deal	duel	dill		dale	dole		doll
s_#	see	sue			say	so		saw
s_p	seep	soup	sip		sapiens	soap	sop	
s_t	seat	suit	sit	soot	sate	(sewed)	sot	sought

(continued)

	/i/	/u/	/ɪ/	/ʊ/	/e/	/o/	/ɑ/	/ɔ/
s_k	seek		sick	forsook	sake	soak	sock	Salk
s_l	seal	(Zuul)	sill		sail	sole		Saul
ʃ_#	she	shoe			shay	show		shaw
ʃ_p	sheep		ship		shape		shop	
ʃ_t	sheet	shoot	shit	(should)	(shade)	(showed)	shot	
ʃ_k	chic		Schick	shook	shake		shock	
ʃ_l			shill		shale	shoal		shawl
k_#	key	coo			kay	(go)		caw
k_p	keep	coop	Kip		cape	cope	cop	
k_t	(Keats)	coot	kit	(could)	Kate	coat	cot	caught
k_k	(geek)	kook	kick	cook	cake	coke	cock	(gawk)
k_l	keel	cool	kill		kale	coal		call
h_#	he	who			hay	hoe	ha	haw
h_p	heap	hoop	hip			hope	hop	
h_t	heat	hoot	hit	(hood)	hate	(hoed)	hot	(hawed)
h_k			hick	hook			hock	hawk
h_l	heel		hill		hale	hole		haul

Appendix B

Wordlists for Chicago Production Experiment

B.1 Wordlist for Production Task

/i/	/u/	/æ/	/o/	/ɑ/	/ɔ/
need	nude	dad	node	nod	gnawed
seed	sued	sad	sewed	sod	sawed
teak	duke	tack	toke	tock	talk
seek	suit	stack	stoke	stock	stalk
dean	dune	dan	don't	don	dawn
heat	hoot	hack	hoax	hock	hawk
geek	kook	cackle	coke	glock	gawk
teed	toot	tat	tote	tot	taught
geese	goose	cat	coat	cot	caught
neat	newt	gnat	note	not	naught
week	wooed	whack	woke	wok	walk
peat	pooch	pad	pose	pod	pawed
beat	boot	bat	boat	bot	bought
feet	food	fat	foes	fodder	fought
peep	poop	pap	pope	pop	pauper
read	root	rat	wrote	rot	wrought
keyed	cooed	cad	code	cod	cawed

(continued)

/i/	/u/	/æ/	/o/	/ɑ/	/ɔ/
eat	ooze	at	oat	otter	ought
sheets	shoots			shots	thoughts
			poked	blocked	balked
				stocks	stalks
				goth	cloth
				gods	frauds
				clocked	caulked

B.2 Phrases for Careful Speech Task

I said **nod** and sod, not **gnawed** and sawed

I said **sod** and tock, not **sawed** and talk

I said **tock** and stock, not **talk** and stalk

I said **stock** and don, not **stalk** and dawn

I said **don** and hock, not **dawn** and hawk

I said **hock** and glock, not **hawk** and gawk

I said **glock** and tot, not **gawk** and taught

I said **tot** and cot, not **taught** and caught

I said **cot** and not, not **caught** and naught

I said **not** and wok, not **naught** and walk

I said **wok** and pod, not **walk** and pawed

I said **pod** and bot, not **pawed** and bought

I said **bot** and fodder, not **bought** and fought

I said **fodder** and pop, not **fought** and pauper

I said **pop** and rot, not **pauper** and wrought

I said **rot** and cod, not **wrought** and cawed

I said **cod** and otter, not **cawed** and ought

I said **otter** and nod, not **ought** and gnawed

I said **gnawed** and sawed, not **nod** and sod

I said **sawed** and talk, not **sod** and tock

I said **talk** and stalk, not **tock** and stock

I said **stalk** and dawn, not **stock** and don

I said **dawn** and hawk, not **don** and hock

I said **hawk** and gawk, not **hock** and glock

I said **gawk** and taught, not **glock** and tot

I said **taught** and caught, not **tot** and cot

I said **caught** and naught, not **cot** and not

I said **naught** and walk, not **not** and wok

I said **walk** and pawed, not **wok** and pod

I said **pawed** and bought, not **pod** and bot

I said **bought** and fought, not **bot** and fodder

I said **fought** and pauper, not **fodder** and pop

I said **pauper** and wrought, not **pop** and rot

I said **wrought** and cawed, not **rot** and cod

I said **cawed** and ought, not **cod** and otter

I said **ought** and gnawed, not **otter** and nod

Appendix C

Stimuli for Chicago Perception Experiment

C.1 Instructions for Perception Task

Please place both hands on the keypad, with your index fingers on the yellow and green keys.

In this task, you will be presented with some made-up words of English. For each new word that is spoken, you will be asked which existing English word it rhymes with. For example, if you hear…

<Practice item 1>

...you will be presented with the following options:

<Practice choices>

Each word will be said in the phrase "say ____ again". The same word may be spoken more than once.

Let's try a few more...

<Practice items 2-5>

Try to answer as quickly and as accurately as possible. If you don't answer in time, the program will automatically advance to the next word.

C.2 Stimuli for Perception Task

nonce word	vowel audio	vowel visual	rhyme 1	rhyme 2
θɑk	/ɑ/	unround	hock	hawk
zɑkt	/ɑ/	unround	blocked	balked
zɑks	/ɑ/	unround	stocks	stalks
skɑθ	/ɑ/	unround	goth	cloth
skɑdz	/ɑ/	unround	gods	frauds
plɑkt	/ɑ/	unround	clocked	caulked
sklɑt	/ɑ/	unround	cot	caught
sklɑts	/ɑ/	unround	shots	thoughts
zɑd	/ɑ/	unround	pod	pawed
zɑt	/ɑ/	unround	bot	bought
θɑk	/ɑ/	round	hock	hawk
zɑkt	/ɑ/	round	blocked	balked
zɑks	/ɑ/	round	stocks	stalks
skɑθ	/ɑ/	round	goth	cloth
skɑdz	/ɑ/	round	gods	frauds
plɑkt	/ɑ/	round	clocked	caulked
sklɑt	/ɑ/	round	cot	caught
sklɑts	/ɑ/	round	shots	thoughts
zɑd	/ɑ/	round	pod	pawed
zɑt	/ɑ/	round	bot	bought
θɔk	/ɔ/	unround	hawk	hock
zɔkt	/ɔ/	unround	balked	blocked
zɔks	/ɔ/	unround	stalks	stocks

(continued)

nonce word	vowel audio	vowel visual	rhyme 1	rhyme 2
skɔθ	/ɔ/	unround	cloth	goth
skɔdz	/ɔ/	unround	frauds	gods
plɔkt	/ɔ/	unround	caulked	clocked
sklɔt	/ɔ/	unround	caught	cot
sklɔts	/ɔ/	unround	thoughts	shots
zɔd	/ɔ/	unround	pawed	pod
zɔt	/ɔ/	unround	bought	bot
θɔk	/ɔ/	round	hawk	hock
zɔkt	/ɔ/	round	balked	blocked
zɔks	/ɔ/	round	stalks	stocks
skɔθ	/ɔ/	round	cloth	goth
skɔdz	/ɔ/	round	frauds	gods
plɔkt	/ɔ/	round	caulked	clocked
sklɔt	/ɔ/	round	caught	cot
sklɔts	/ɔ/	round	thoughts	shots
zɔd	/ɔ/	round	pawed	pod
zɔt	/ɔ/	round	bought	bot
θek	/e/	mid	fake	cheek
zekt	/e/	mid	baked	peeked
zeks	/e/	mid	stakes	speaks
skeθ	/e/	mid	eighth	heath
skedz	/e/	mid	fades	needs
plekt	/e/	mid	ached	creaked

(continued)

nonce word	vowel audio	vowel visual	rhyme 1	rhyme 2
sklet	/e/	mid	gate	heat
sklets	/e/	mid	states	sheets
zed	/e/	mid	payed	peed
zet	/e/	mid	hate	beat
θek	/e/	high	fake	cheek
zekt	/e/	high	baked	peeked
zeks	/e/	high	stakes	speaks
skeθ	/e/	high	eighth	heath
skedz	/e/	high	fades	needs
plekt	/e/	high	ached	creaked
sklet	/e/	high	gate	heat
sklets	/e/	high	states	sheets
zed	/e/	high	payed	peed
zet	/e/	high	hate	beat
θok	/o/	mid	folk	fake
zokt	/o/	mid	poked	baked
zoks	/o/	mid	stokes	stakes
skoθ	/o/	mid	oath	eighth
skodz	/o/	mid	nodes	fades
plokt	/o/	mid	cloaked	ached
sklot	/o/	mid	coat	gate
sklots	/o/	mid	floats	states
zod	/o/	mid	bode	payed

(continued)

nonce word	vowel audio	vowel visual	rhyme 1	rhyme 2
zot	/o/	mid	bloat	hate
θok	/o/	high	folk	fluke
zokt	/o/	high	poked	spooked
zoks	/o/	high	stokes	dukes
skoθ	/o/	high	oath	tooth
skodz	/o/	high	nodes	nudes
plokt	/o/	high	cloaked	nuked
sklot	/o/	high	coat	hoot
sklots	/o/	high	floats	chutes
zod	/o/	high	bode	prude
zot	/o/	high	bloat	shoot
θik	/i/	high	cheek	fluke
zikt	/i/	high	peeked	spooked
ziks	/i/	high	speaks	dukes
skiθ	/i/	high	heath	tooth
skidz	/i/	high	needs	nudes
plikt	/i/	high	creaked	nuked
sklit	/i/	high	heat	hoot
sklits	/i/	high	sheets	chutes
zid	/i/	high	peed	prude
zit	/i/	high	beat	shoot
θik	/i/	mid	cheek	fake
zikt	/i/	mid	peeked	baked

(continued)

nonce word	vowel audio	vowel visual	rhyme 1	rhyme 2
ziks	/i/	mid	speaks	stakes
skiθ	/i/	mid	heath	eighth
skidz	/i/	mid	needs	fades
plikt	/i/	mid	creaked	ached
sklit	/i/	mid	heat	gate
sklits	/i/	mid	sheets	states
zid	/i/	mid	peed	payed
zit	/i/	mid	beat	hate
θuk	/u/	high	fluke	cheek
zukt	/u/	high	spooked	peeked
zuks	/u/	high	dukes	speaks
skuθ	/u/	high	tooth	heath
skudz	/u/	high	nudes	needs
plukt	/u/	high	nuked	creaked
sklut	/u/	high	hoot	heat
skluts	/u/	high	chutes	sheets
zud	/u/	high	prude	peed
zut	/u/	high	shoot	beat
θuk	/u/	mid	fluke	folk
zukt	/u/	mid	spooked	poked
zuks	/u/	mid	dukes	stokes
skuθ	/u/	mid	tooth	oath
skudz	/u/	mid	nudes	nodes

(continued)

nonce word	vowel audio	vowel visual	rhyme 1	rhyme 2
plukt	/u/	mid	nuked	cloaked
sklut	/u/	mid	hoot	coat
skluts	/u/	mid	chutes	floats
zud	/u/	mid	prude	bode
zut	/u/	mid	shoot	bloat

References

Adank, Patti, Roel Smits, and Roeland van Hout. 2004. "A Comparison of Vowel Normalization Procedures for Language Variation Research." *The Journal of the Acoustical Society of America* 116 (5): 3099–3107. doi:10.1121/1.1795335.

Anderson, Stephen R. 1976. "On the Description of Multiply-Articulated Consonants." *Journal of Phonetics* 4 (1): 17–27.

Articulate Instruments Ltd. 2008. *Ultrasound Stabilisation Headset Users Manual: Revision 1.4.* Edinburgh, UK: Articulate Instruments Ltd.

———. 2012. *Articulate Assistant Advanced User Guide: Version 2.14.* Edinburgh, UK: Articulate Instruments Ltd.

Baker, Adam. 2006. "Quantifying Diphthongs: A Statistical Technique for Distinguishing Formant Contours." Paper presented at NWAV 35, Columbus, OH.

Baker, Adam, Diana Archangeli, and Jeff Mielke. 2011. "Variability in American English S-Retraction Suggests a Solution to the Actuation Problem." *Language Variation and Change* 23 (3): 347–374. doi:10.1017/S0954394511000135.

Baker, Peter S. 2012. *Introduction to Old English.* 3rd ed. Malden, MA: Wiley-Blackwell.

Bakst, Sarah, and Susan Lin. 2015. "An Ultrasound Investigation Into Articulatory Variation in American /r/ and /s/." In *Proceedings of the 18th International Congress of Phonetic Sciences,* edited by The Scottish Consortium for ICPhS 2015. Glasgow, UK: The University of Glasgow.

Baranowski, Maciej A. 2006. “Phonological Variation and Change in the Dialect of Charleston, South Carolina.” Doctoral dissertation, University of Pennsylvania. http://search.proquest.com/docview/305246681.

———. 2008. “The Fronting of the Back Upgliding Vowels in Charleston, South Carolina.” *Language Variation and Change* 20 (3): 527–551. doi:10.1017/S0954394508000136.

Bateman, Nicoleta. 2007. “A Crosslinguistic Investigation of Palatalization.” Doctoral dissertation, University of California, San Diego. http://escholarship.org/uc/item/13s331md.

———. 2010. “The Change from Labial to Palatal as Glide Hardening.” *Linguistic Typology* 14 (2-3): 167–211. doi:10.1515/lity.2010.008.

———. 2011. “On the Typology of Palatalization.” *Language and Linguistics Compass* 5 (8): 588–602. doi:10.1111/j.1749-818X.2011.00294.x.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. doi:10.18637/jss.v067.i01.

Baudouin de Courtenay, Jan. 1895. *A Baudouin De Courtenay Anthology: The Beginnings of Structural Linguistics.* Edited by Edward Stankiewicz. Bloomington, IN: Indiana University Press.

Beddor, Patrice Speeter, Rena Arens Krakow, and Louis M. Goldstein. 1986. “Perceptual Constraints and Phonological Change: A Study of Nasal Vowel Height.” *Phonology Yearbook* 3:197–217. doi:10.1017/S0952675700000646.

Bhat, D.N.S. 1978. “A General Study of Palatalization.” In *Phonology,* vol. 2 of *Universals of Human Language,* edited by Joseph H. Greenberg, 47–92. Stanford, CA: Stanford University Press.

Bladon, R. Anthony W., and Francis J. Nolan. 1977. “A Video-Fluorographic Investigation of Tip and Blade Alveolars in English.” *Journal of Phonetics* 5:185–193.

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns.* Cambridge: Cambridge University Press.

———. 2006. “A Theoretical Synopsis of Evolutionary Phonology.” *Theoretical Linguistics* 32 (2): 117–166. doi:10.1515/TL.2006.009.

Blevins, Juliette, and Andrew Garrett. 2004. “The Evolution of Metathesis.” In *Phonetically-Based Phonology,* edited by Bruce Hayes, Robert Kirchner, and Donca Steriade, 117–156. Cambridge: Cambridge University Press.

Boersma, Paul, and David Weenink. 2017. *Praat: Doing Phonetics by Computer (version 6.0.36).* www.praat.org.

Browman, Catherine P., and Louis Goldstein. 1989. “Articulatory Gestures as Phonological Units.” *Phonology* 6 (2): 201–251. doi:10.1017/S0952675700001019.

Brunner, Jana, Susanne Fuchs, and Pascal Perrier. 2009. “On the Relationship Between Palate Shape and Articulatory Behavior.” *The Journal of the Acoustical Society of America* 125 (6): 3936–49. doi:10.1121/1.3125313.

Buschmeier, Hendrik, and Marcin Włodarczak. 2013. “TextGridTools: A TextGrid Processing and Analysis Toolkit for Python.” In *Proceedings der 27. Konferenz zur Elektronischen Sprachsignalverarbeitung,* 152–157.

Bybee, Joan. 2000. “The Phonology of the Lexicon: Evidence from Lexical Diffusion.” In *Usage-Based Models of Language,* edited by Michael Barlow and Suzanne Kemmer, 65–85. Stanford: CSLI.

Cahill, Michael. 1999. “Aspects of the Phonology of Labial-Velar Stops.” *Studies in African Linguistics* 28 (2): 155–184.

Calabrese, Andrea. 2000. “The Feature [ATR] and Vowel Fronting in Romance.” In *Phonological Theory and the Dialects of Italy,* edited by Lori Repetti, 59–88. Amsterdam: John Benjamins.

Chang, Steve, Madelaine C. Plauché, and John J. Ohala. 2001. “Markedness and Consonant Confusion Asymmetries.” In *The Role of Speech Perception in Phonology,* edited by Elizabeth Hume and Keith Johnson, 79–101. London: Academic Press.

Chen, Yu, and Hua Lin. 2011. “Analysing Tongue Shape and Movement in Vowel Production Using SS ANOVA in Ultrasound Imaging.” In *Proceedings of the 17th International Congress of Phonetic Sciences,* edited by Wai Sum Lee and Eric Zee, 124–127. Hong Kong: City University of Hong Kong.

Childers, Donald G. 1978. *Modern Spectrum Analysis.* IEEE Computer Society Press.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper & Row.

Cole, Desmond T. 1955. *An Introduction to Tswana Grammar.* London: Longmans, Green, and Co.

Cooper, Franklin S., Pierre Delattre, Alvin M. Liberman, John M. Borst, and Louis J. Gerstman. 1952. “Some Experiments on the Perception of Synthetic Speech Sounds.” *The Journal of the Acoustical Society of America* 24 (6): 597–606. doi:10.1121/1.1906940.

Cox, Felicity. 1999. “Vowel Change in Australian English.” *Phonetica* 56 (1-2): 1–27. doi:10.1159/000028438.

Cox, Felicity, and Sallyanne Palethorpe. 2001. “The Changing Face of Australian English Vowels.” In *English in Australia,* edited by David Blair and Peter Collins, 17–44. Varieties of English Around the World 26. Amsterdam: John Benjamins.

Davidson, Lisa. 2006. “Comparing Tongue Shapes from Ultrasound Imaging Using Smoothing Spline Analysis of Variance.” *The Journal of the Acoustical Society of America* 120 (1): 407–415. doi:10.1121/1.2205133.

De Decker, Paul M., and Jennifer Nycz. 2012. “Are Tense [æ]s Really Tense? The Mapping Between Articulation and Acoustics.” *Lingua* 122 (7): 810–821. doi:10.1016/j.lingua.2012.01.003.

De Boer, Bart. 2001. *The Origins of Vowel Systems.* Oxford: Oxford University Press.

De Jong, Kenneth J. 1994. “On the Status of Redundant Features: The Case of Backing and Rounding in American English.” *Working Papers of the Cornell Phonetics Laboratory* 9:93–114.

Delattre, Pierre, and Donald C. Freeman. 1968. “A Dialect Study of American R’s by X-Ray Motion Picture.” *Linguistics* 6 (44): 29–68. doi:10.1515/ling.1968.6.44.29.

Diehl, Randy L., and Keith R. Kluender. 1989. “On the Objects of Speech Perception.” *Ecological Psychology* 1 (2): 121–144.

Dinkin, Aaron J. 2009. “Dialect Boundaries and Phonological Change in Upstate New York.” Doctoral dissertation, University of Pennsylvania.

Driscoll, Anna, and Emma Lape. 2015. “Reversal of the Northern Cities Shift in Syracuse, New York.” *University of Pennsylvania Working Papers in Linguistics* 21 (2).

Eckert, Penelope. 1989. *Jocks and Burnouts: Social Categories and Identity in the High School.* New York: Teachers College Press.

———. 2008. “Where Do Ethnolects Stop?” *International Journal of Bilingualism* 12 (1-2): 25–42. doi:10.1177/13670069080120010301.

Espy-Wilson, Carol Y. 2004. “Articulatory Strategies, Speech Acoustics and Variability.” In *Proceedings of Sound to Sense: Fifty+ Years of Discoveries in Speech Communication,* edited by Janet Slifka, Sharon Manuel, and Melanie Matthies, B62–B76. Cambridge, MA: MIT Research Laboratory of Electronics.

Fant, Gunnar. 1960. *Acoustic Theory of Speech Production.* The Hague: Mouton.

———. 1973. *Speech Sounds and Features.* Cambridge, MA: MIT Press.

Fasold, Ralph W. 1969. “A Sociolinguistic Study of the Pronunciation of Three Vowels in Detroit Speech.” Unpublished Ms.

Ferguson, Sarah Hargus. 2004. “Talker Differences in Clear and Conversational Speech: Vowel Intelligibility for Normal-Hearing Listeners.” *The Journal of the Acoustical Society of America* 116 (4): 2365–2373.

FFmpeg Developers. 2018. *FFmpeg v3.4.2 [Computer Software].* https://ffmpeg.org.

Flemming, Edward. 2004. “Contrast and Perceptual Distinctiveness.” In *Phonetically-Based Phonology,* edited by Bruce Hayes, Robert Kirchner, and Donca Steriade. Cambridge: Cambridge University Press.

Foulkes, Paul. 1997. “Historical Laboratory Phonology: Investigating /p/ > /f/ > /h/ Changes.” *Language and Speech* 40 (3): 249–276.

Foulkes, Paul, and Gerard J. Docherty. 2000. “Another Chapter in the Story of /r/: ‘Labiodental’ Variants in British English.” *Journal of Sociolinguistics* 4 (1): 30–59. doi:10.1111/1467-9481.00102.

Fowler, Carol A. 1986. “An Event Approach to the Study of Speech Perception from a Direct-Realist Perspective.” *Journal of Phonetics* 14:3–28.

———. 1996. “Listeners Do Hear Sounds, Not Tongues.” *The Journal of the Acoustical Society of America* 99 (3): 1730–1741. doi:10.1121/1.415237.

Fowler, Carol A., and Dawn J. Dekle. 1991. “Listening with Eye and Hand: Cross-Modal Contributions to Speech Perception.” *Journal of Experimental Psychology: Human Perception and Performance* 17 (3): 816–828. doi:10.1037/0096-1523.17.3.816.

Friedman, Lauren. 2014. “The St. Louis Corridor: Mixing, Competing, and Retreating Dialects.” Doctoral dissertation, University of Pennsylvania.

Fruehwald, Josef. 2010. “SS ANOVA.” Unpublished Ms.

Gagné, Jean Pierre, Valerie Masterson, Kevin G. Munhall, Nancy Bilida, and Carol Querengesser. 1994. “Across Talker Variability in Auditory, Visual, and Audiovisual Speech Intelligibility for Conversational and Clear Speech.” *Journal of the Academy of Rehabilitative Audiology* 27:135–158.

Galantucci, Bruno, Carol A. Fowler, and M. T. Turvey. 2006. “The Motor Theory of Speech Perception Reviewed.” *Psychonomic Bulletin & Review* 13 (3): 361–377. doi:10.3758/bf03193857.

Garrett, Andrew, and Keith Johnson. 2011. “Phonetic Bias in Sound Change.” In *UC Berkeley Phonology Lab Annual Report,* 9–61. Berkeley: University of California, Berkeley.

Gick, Bryan, and Donald Derrick. 2009. “Aero-Tactile Integration in Speech Perception.” *Nature* 462 (7272): 502–504. doi:10.1038/nature08572.

Gilmore, Grover C., H. Hersh, A. Caramazza, and J. Griffin. 1979. “Multidimensional Letter Similarity Derived from Recognition Errors.” *Perception & Psychophysics* 25 (5): 425–431. doi:10.3758/BF03199852.

Gluth, Caroline, and Philip Hoole. 2015. “How Can Speech Production Skills Be Predicted from Visual, Auditory and Haptic Perception Skills?” In *Proceedings of the 18th International Congress of Phonetic Sciences,* edited by The Scottish Consortium for ICPhS 2015. Glasgow: The University of Glasgow.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury, and Peter Trudgill. 2004. *New Zealand English: Its Origins and Evolution.* Cambridge: Cambridge University Press.

Gordon, Matthew J. 1997. “Urban Sound Change Beyond City Limits: The Spread of the Northern Cities Shift in Michigan.” Doctoral dissertation, University of Michigan. http://hdl.handle.net/2027.42/130716.

Gu, Chong. 2002. *Smoothing Spline ANOVA Models.* New York: Springer.

Guion, Susan. 1998. “The Role of Perception in the Sound Change of Velar Palatalization.” *Phonetica* 55 (1-2): 18–52. doi:10.1159/000028423.

Guthrie, Malcolm. 1970. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages.* Farnborough, UK: Gregg Press.

Hagiwara, Robert. 1995. “Acoustic Realizations of American /r/ as Produced by Women and Men.” *UCLA Working Papers in Phonetics* 90. http://www.escholarship.org/uc/item/8779b7gq.

———. 1997. “Dialect Variation and Formant Frequency: The American English Vowels Revisited.” *The Journal of the Acoustical Society of America* 102 (1): 655–658. doi:10.1121/1.419712.

Hall-Lew, Lauren. 2009. “Ethnicity and Phonetic Variation in a San Francisco Neighborhood.” Doctoral dissertation, Stanford University. http://search.proquest.com/docview/304997888.

Harrington, Jonathan, Felicitas Kleber, and Ulrich Reubold. 2008. “Compensation for Coarticulation, /u/-Fronting, and Sound Change in Standard Southern British: An Acoustic and Perceptual Study.” *The Journal of the Acoustical Society of America* 123 (5): 2825–2835. doi:10.1121/1.2897042.

———. 2011. “The Contributions of the Lips and the Tongue to the Diachronic Fronting of High Back Vowels in Standard Southern British English.” *Journal of the International Phonetic Association* 41 (2): 137–156. doi:10.1017/S0025100310000265.

Havenhill, Jonathan, and Youngah Do. 2018. “Visual Speech Perception Cues Constrain Patterns of Articulatory Variation and Sound Change.” *Frontiers in Psychology* 9:728. doi:10.3389/fpsyg.2018.00728.

Hay, Jennifer, Aaron Nolan, and Katie Drager. 2006. “From Fush to Feesh: Exemplar Priming in Speech Perception.” *The Linguistic Review* 23 (3): 351–379. doi:10.1515/TLR.2006.014.

Hayes, Bruce. 1997. “Phonetically Driven Phonology: The Role of Optimality Theory and Inductive Grounding.” In *Optimality Theory in Phonology: A Reader,* edited by John J. McCarthy, 290–309. Malden, MA: Blackwell. doi:10.1002/9780470756171.ch15.

Hayes, Bruce, Robert Kirchner, and Donca Steriade, eds. 2004. *Phonetically Based Phonology.* Cambridge: Cambridge University Press.

Hickey, Raymond. 1984. “On the Nature of Labial Velar Shift.” *Journal of Phonetics* 12:345–354.

Hillenbrand, James, Laura A. Getty, Kimberlee Wheeler, and Michael J. Clark. 1994. “Acoustic Characteristics of American English Vowels.” *The Journal of the Acoustical Society of America* 95 (5): 2875–2875. doi:10.1121/1.409456.

Hinton, Leanne, Birch Moonwomon, Sue Bremner, Herb Luthin, Mary Van Clay, Jean Lerner, and Hazel Corcoran. 1987. “It’s Not Just the Valley Girls: A Study of California English.” *Annual Meeting of the Berkeley Linguistics Society* 13:117–128. doi:10.3765/bls.v13i0.1811.

Hooper, Joan B. 1976. “Word Frequency in Lexical Diffusion and the Source of Morphophonological Change.” In *Current Progress in Historical Linguistics: Proceedings of the Second International Conference on Historical Linguistics,* edited by William M. Christie, 96–105. Amsterdam: North Holland.

Houde, John F., and Michael I. Jordan. 1998. “Sensorimotor Adaptation in Speech Production.” *Science* 279 (5354): 1213–1216. doi:10.1126/science.279.5354.1213.

Hyman, Larry M. 1975. *Phonology: Theory and Analysis.* New York: Holt, Rinehart, & Winston.

Ito, Rika. 1999. “Diffusion of Urban Sound Change in Rural Michigan: A Case of the Northern Cities Shift.” Doctoral dissertation, Michigan State University. http://search.proquest.com/docview/304531466.

Jakobson, Roman, Gunnar Fant, and Morris Halle. 1951. *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates.* Cambridge, MA: MIT Press.

Johnson, Keith. 1997. “Speech Perception Without Speaker Normalization: An Exemplar Model.” In *Talker Variation in Speech Processing,* edited by Keith Johnson and John W. Mullennix, 145–165. San Diego: Academic Press.

———. 2015. “Audio-Visual Factors in Stop Debuccalization in Consonant Sequences.” In *UC Berkely Phonology Lab Annual Report,* 227–242. Berkeley: University of California, Berkeley. http://escholarship.org/uc/item/6hc5k2zq.

Johnson, Keith, Christian T. DiCanio, and Laurel MacKenzie. 2007. “The Acoustic and Visual Phonetic Basis of Place of Articulation in Excrescent Nasals.” In *UC Berkeley Phonology Lab Annual Report,* 529–561. Berkeley: University of California, Berkeley. http://escholarship.org/uc/item/0n17f74x.

Johnson, Keith, Edward Flemming, and Richard Wright. 1993. “The Hyperspace Effect: Phonetic Targets Are Hyperarticulated.” *Language* 69 (3): 505–528.

Jones, Jeffery A., and Kevin G. Munhall. 2005. “Remapping Auditory-Motor Representations in Voice Production.” *Current Biology* 15 (19): 1768–1772. doi:10.1016/j.cub.2005.08.063.

Karlgren, Bernhard. 1915. *Études sur la phonologie chinoise.* Uppsala, Sweden: K.W. Appelberg.

Kendall, Tyler, and Erik R. Thomas. 2014. *Vowels: Vowel Manipulation, Normalization, and Plotting*. R package version 1.2-1. https://CRAN.R-project.org/package=vowels.

Kenstowicz, Michael, and Charles Kisseberth. 1977. *Topics in Phonological Theory.* New York: Academic Press.

Kirby, James P. 2011. "Vietnamese (Hanoi Vietnamese)." *Journal of the International Phonetic Association* 41 (3): 381–392.

Kochetov, Alexei. 1998. "Labial Palatalization: A Gestural Account of Phonetic Implementation." *The Canadian Linguistic Association Annual Proceedings:* 38–50.

———. 2011. "Palatalization." In *The Blackwell Companion to Phonology,* edited by Colin Ewen, Beth Hume, Marc van Oostendorp, and Keren Rice, 1666–1690. Malden, MA: Wiley-Blackwell.

———. 2016. "Palatalization and Glide Strengthening as Competing Repair Strategies: Evidence from Kirundi." *Glossa* 1 (1): 1–31. doi:10.5334/gjgl.32.

Krakow, Rena A., Patrice Speeter Beddor, Louis M. Goldstein, and Carol A. Fowler. 1988. "Coarticulatory Influences on the Perceived Height of Nasal Vowels." *The Journal of the Acoustical Society of America* 83 (3): 1146–1158. doi:10.1121/1.396059.

Kricos, Patricia B. 1996. "Differences in Visual Intelligibility Across Talkers." In *Speechreading by Humans and Machines: Models, Systems, and Applications,* edited by David G. Stork and Marcus E. Hennecke, 43–53. Berlin: Springer. doi:10.1007/978-3-662-13015-5_4.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82 (13): 1–26. doi:10.18637/jss.v082.i13.

Labov, William. 1963. “The Social Motivation of a Sound Change.” *Word* 19:273–309. doi:10.1080/00437956.1963.11659799.

———. 1991. “Three Dialects of English.” In *New Ways of Analyzing Variation in English,* edited by Penelope Eckert, 1–4. New York: Academic Press.

———. 1994. *Principles of Linguistic Change.* Malden, MA: Wiley-Blackwell.

———. 2007. “Transmission and Diffusion.” *Language* 83 (2): 344–387.

Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English.* Berlin: Walter de Gruyter. doi:10.1515/9783110206838.

Labov, William, Mark Karen, and Corey Miller. 1991. “Near-Mergers and the Suspension of Phonemic Contrast.” *Language Variation and Change* 3 (1): 33–74.

Labov, William, Malcah Yaeger, and Richard Steiner. 1972. *A Quantitative Study of Sound Change in Progress.* Philadelphia: The US Regional Survey.

Ladefoged, Peter, Joseph DeClerk, Mona Lindau, and George Papçun. 1972. “An Auditory-Motor Theory of Speech Production.” *UCLA Working Papers in Phonetics* 22:48–76.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World’s Languages.* Oxford: Blackwell.

Lane, Harlan, Melanie Matthies, Joseph Perkell, Vick Jennell, and Majid Zandipour. 2001. “The Effects of Changes in Hearing Status in Cochlear Implant Users on the Acoustic Vowel Space and CV Coarticulation.” *Journal of Speech, Language, and Hearing Research* 44 (3): 552–63. doi:10.1044/1092-4388(2001/043).

Lawson, Eleanor, James M. Scobbie, and Jane Stuart-Smith. 2011. "The Social Stratification of Tongue Shape for Postvocalic /r/ in Scottish English." *Journal of Sociolinguistics* 15 (2): 256–268. doi:10.1111/j.1467-9841.2011.00464.x.

Lawson, Eleanor, Jane Stuart-Smith, and Lydia Mills. 2017. *Using Ultrasound to Investigate Articulatory Variation in the GOOSE Vowel in the British Isles.* Paper presented at Ultrafest VIII, Potsdam, Germany.

Lawson, Eleanor, Jane Stuart-Smith, James M. Scobbie, Malcah Yaeger-Dror, and Margaret Maclagan. 2010. "Liquids." In *Sociophonetics: A Student's Guide,* edited by Marianna Di Paolo and Malcah Yaeger-Dror, 73–86. New York: Routledge.

Lee-Kim, Sang Im, Lisa Davidson, and Sangjin Hwang. 2013. "Morphological Effects on the Darkness of English Intervocalic /l/." *Laboratory Phonology* 4 (2): 475–511. doi:10.1515/lp-2013-0015.

Lee-Kim, Sang Im, Shigeto Kawahara, and Seunghun J. Lee. 2014. "The 'Whistled' Fricative in Xitsonga: Its Articulation and Acoustics." *Phonetica* 71 (1): 50–81. doi:10.1159/000362672.

Lemle, Miriam. 1971. "Internal Classification of the Tupí-Guaraní Linguistic Family." In *Tupí Studies I,* edited by David Bendor-Samuel, 107–129. Publications in Linguistics and Related Fields 29. Norman, OK: Summer Institute of Linguistics.

Liberman, Alvin M. 1957. "Some Results of Research on Speech Perception." *The Journal of the Acoustical Society of America* 29 (1): 117–123. doi:10.1121/1.1908635.

Liberman, Alvin M., Franklin S. Cooper, Donald P. Shankweiler, and Michael Studdert-Kennedy. 1967. "Perception of the Speech Code." *Psychological Review* 74 (6): 431–461. doi:10.1037/h0020279.

Liberman, Alvin M., Pierre C. Delattre, and Franklin S. Cooper. 1952. “The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants.” *The American Journal of Psychology* 65 (4): 497–516. doi:10.2307/1418032.

Liberman, Alvin M., and Ignatius G. Mattingly. 1985. “The Motor Theory of Speech Perception Revised.” *Cognition* 21 (1): 1–36.

Liljencrants, Johan, and Björn Lindblom. 1972. “Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast.” *Language* 48 (4): 839–862.

Lindblom, Björn. 1986. “Phonetic Universals in Vowel Systems.” In *Experimental Phonology,* edited by John J. Ohala and Jeri J. Jaeger, 13–44. Orlando: Academic Press.

———. 1990. “Explaining Phonetic Variation: A Sketch of the H&H Theory.” In *Speech Production and Speech Modelling,* edited by William J. Hardcastle and Alain Marchal, 403–439. Dordrecht: Springer. doi:10.1007/978-94-009-2037-8_16.

Lindblom, Björn, Susan Guion, Susan Hura, Seung-Jae Moon, and Raquel Willerman. 1995. “Is Sound Change Adaptive?” *Rivista di Linguistica* 7:5–36.

Lisker, Leigh. 1978. “Rapid vs Rabid: A Catalogue of Acoustical Features That May Cue the Distinction.” In *Haskins Laboratories Status Report on Speech Research,* 54:127–132. New Haven, CT: Haskins Laboratories.

———. 1986. “‘Voicing’ in English: A Catalogue of Acoustic Features Signaling /b/ Versus /p/ in Trochees.” *Language and Speech* 29:3–11.

Lyublinskaya, V.V. 1966. “Recognition of Articulation Cues in Stop Consonants in Transition from Vowel to Consonant.” *Soviet Physics-Acoustics* 12 (2): 185–192.

Maddieson, Ian. 1984. *Patterns of Sounds.* Cambridge: Cambridge University Press.

Maddieson, Ian, and Kristin Precoda. 1990. “Updating UPSID.” *UCLA Working Papers in Phonetics* 74:104–111.

Majors, Tivoli, and Matthew J. Gordon. 2008. “The [+spread] of the Northern Cities Shift.” *University of Pennsylvania Working Papers in Linguistics* 14 (2): 111–120. http://repository.upenn.edu/pwpl/vol14/iss2/14.

Mayer, Connor, Bryan Gick, Weigel Tamra, and Doug H. Whalen. 2013. “Perceptual Integration of Visual Evidence of the Airstream from Aspirated Stops.” *Canadian Acoustics* 41 (3): 23–27. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4474184/.

Mazzaro, Natalia. 2010. “Changing Perceptions: The Sociophonetic Motivations of the Labial Velar Alternation in Spanish.” In *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology,* edited by Marta Ortega-Llebaria, 128–145. Somerville, MA: Cascadilla Proceedings Project.

McCarthy, Corrine. 2010. “The Northern Cities Shift in Real Time: Evidence from Chicago.” *University of Pennsylvania Working Papers in Linguistics* 15 (2).

McGuire, Grant, and Molly Babel. 2012. “A Cross-Modal Account for Synchronic and Diachronic Patterns of /f/ and /θ/ in English.” *Laboratory Phonology* 3 (2): 1–41. doi:10.1515/lp-2012-0014.

McGurk, Harry, and John MacDonald. 1976. “Hearing Lips and Seeing Voices.” *Nature* 264:746–748. doi:10.1038/264746a0.

Ménard, Lucie, Sophie Dupont, Shari R. Baum, and Jérôme Aubin. 2009. “Production and Perception of French Vowels by Congenitally Blind Adults and Sighted Adults.” *The Journal of the Acoustical Society of America* 126 (3): 1406–1414. doi:10.1121/1.3158930.

Ménard, Lucie, Corinne Toupin, Shari R. Baum, Serge Drouin, Jérôme Aubin, and Mark Tiede. 2013. “Acoustic and Articulatory Analysis of French Vowels Produced by Congenitally Blind Adults and Sighted Adults.” *The Journal of the Acoustical Society of America* 134 (4): 2975–87. doi:10.1121/1.4818740.

Ménard, Lucie, Paméla Trudeau-Fisette, Dominique Côté, Marie Bellavance-Courtemanche, and Christine Turgeon. 2015. “Acoustic and Articulatory Correlates of Speaking Condition in Blind and Sighted Speakers.” In *Proceedings of the 18th International Congress of Phonetic Sciences,* edited by The Scottish Consortium for ICPhS 2015. Glasgow: The University of Glasgow.

Ménard, Lucie, Pamela Trudeau-Fisette, Dominique Côté, and Christine Turgeon. 2016. “Speaking Clearly for the Blind: Acoustic and Articulatory Correlates of Speaking Conditions in Sighted and Congenitally Blind Speakers.” *PLOS ONE* 11 (9): e0160088. doi:10.1371/journal.pone.0160088.

Mesthrie, Rajend. 2010. “Socio-Phonetics and Social Change: Deracialisation of the GOOSE Vowel in South African English.” *Journal of Sociolinguistics* 14 (1): 3–33.

Mielke, Jeff. 2015. “An Ultrasound Study of Canadian French Rhotic Vowels with Polar Smoothing Spline Comparisons.” *The Journal of the Acoustical Society of America* 137 (5): 2858–2869. doi:10.1121/1.4919346.

Mielke, Jeff, Adam Baker, and Diana Archangeli. 2010. “Variability and Homogeneity in American English /ɹ/ Allophony and /s/ Retraction.” In *Laboratory Phonology 10,* edited by Cécile Fougeron, Barbara Kuehnert, Mariapaola Imperio, and Nathalie Vallee, 699–729. Berlin: Mouton de Gruyter.

———. 2016. “Individual-Level Contact Limits Phonological Complexity: Evidence from Bunched and Retroflex /ɹ/.” *Language* 92 (1): 101–140.

Mines, M. Ardussi, Barbara F. Hanson, and June E. Shoup. 1978. “Frequency of Occurrence of Phonemes in Conversational English.” *Language and Speech* 21 (3): 221–241.

Minkova, Donka. 2004. “Philology, Linguistics, and the History of [hw]~[w].” In *Studies in the History of the English Language II: Unfolding Conversations,* edited by Anne Curzan and Kimberly Emmons. Topics in English Linguistics 45. Berlin: Mouton de Gruyter.

Mutaka, Ngessimo M., and Carl Ebobissé. 1996. “The Formation of Labial-Velars in Sawabantu: Evidence for Feature Geometry.” *Journal of West African Languages* 26:3–14.

Nasir, Sazzad M., and David J. Ostry. 2006. “Somatosensory Precision in Speech Production.” *Current Biology* 16 (19): 1918–1923. doi:10.1016/j.cub.2006.07.069.

Nearey, Terrance Michael. 1978. “Phonetic Feature Systems for Vowels.” Doctoral dissertation, University of Alberta.

Noiray, Aude, Khalil Iskarous, and Douglas H. Whalen. 2014. “Variability in English Vowels is Comparable in Articulation and Acoustics.” *Laboratory Phonology* 5 (2): 271–288.

Nycz, Jennifer, and Paul De Decker. 2006. “A New Way of Analyzing Vowels: Comparing Formant Contours Using Smoothing Spline ANOVA.” Poster presented at NWAV 35, Columbus, OH.

O’Brien, Jeremy Paul. 2012. “An Experimental Approach to Debuccalization and Supplementary Gestures.” Doctoral dissertation, UC Santa Cruz.

OED Online. 2016. Oxford: Oxford University Press. Accessed January 20, 2016. http://www.oed.com.

Ohala, John J. 1975. "Phonetic Explanations for Nasal Sound Patterns." In *Nasálfest: Papers from a Symposium on Nasals and Nasalization,* edited by Charles A. Ferguson, Larry M. Hyman, and John J. Ohala, 289–316. Stanford, CA: Language Universals Project.

———. 1978. "Southern Bantu vs. the World: The Case of Palatalization of Labials." *Annual Meeting of the Berkeley Linguistics Society* 4:370–386. doi:10.3765/bls.v4i0.2218.

———. 1981. "The Listener as a Source of Sound Change." In *Papers from the Parasession on Language and Behavior,* edited by Carrie S. Masek, Roberta A. Hendrick, and Mary F. Miller, 178–203. Chicago: Chicago Linguistic Society.

———. 1983. "The Origin of Sound Patterns in Vocal Tract Constraints." In *The Production of Speech,* edited by Peter F. MacNeilage, 189–216. New York: Springer.

———. 1989. "Sound Change is Drawn from a Pool of Synchronic Variation." In *Language Change: Contributions to the Study of its Causes,* edited by Leiv E. Breivik and Ernst H. Jahr, 173–198. Berlin: Mouton de Gruyter.

———. 1993. "The Phonetics of Sound Change." In *Historical Linguistics: Problems and Perspectives,* edited by Charles Jones, 237–278. London: Longman.

Ohala, John J., and James Lorentz. 1977. "The Story of [w]: An Exercise in the Phonetic Explanation for Sound Patterns." *Proceedings of the Annual Meeting of the Berkeley Linguistic Society* 3:577–599.

Paul, Hermann. 1890. *Principles of the History of Language.* London: Longmans, Green, and Co.

Peirce, Jonathan W. 2007. "PsychoPy: Psychophysics Software in Python." *Journal of Neuroscience Methods* 162 (1-2): 8–13. doi:10.1016/j.jneumeth.2006.11.017.

Perkell, Joseph S., Melanie L. Matthies, Mario A. Svirsky, and Michael I. Jordan. 1993. "Trading Relations Between Tongue-Body Raising and Lip Rounding in Production of the Vowel /u/: A Pilot 'Motor Equivalence' Study." *The Journal of the Acoustical Society of America* 93 (5): 2948–2961. doi:10.1121/1.405814.

Peterson, Gordon E., and Harold L. Barney. 1952. "Control Methods Used in a Study of the Vowels." *The Journal of the Acoustical Society of America* 24 (2): 175–184. doi:10.1121/1.1906875.

Picheny, M. A., N. I. Durlach, and L. D. Braida. 1986. "Speaking Clearly for the Hard of Hearing Ii." *Journal of Speech, Language, and Hearing Research* 29 (4): 434–446. doi:10.1044/jshr.2904.434.

Pierrehumbert, Janet B. 2001. "Exemplar Dynamics: Word Frequency, Lenition and Contrast." In *Typological Studies in Language,* edited by Joan Bybee and Paul Hopper, 137–157. Amsterdam: John Benjamins. doi:10.1075/tsl.45.08pie.

Podesva, Robert J. 2011. "The California Vowel Shift and Gay Identity." *American Speech* 86 (1): 32–51. doi:10.1215/00031283-1277501.

Pratt, Teresa, and Annette D'Onofrio. 2017. "Jaw Setting and the California Vowel Shift in Parodic Performance." *Language in Society* 46 (3): 283–312.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. 1992. *Numerical Recipes in C.* 2nd ed. Cambridge: Cambridge University Press.

Preston, Dennis R. 1986. “Five Visions of America.” *Language in Society* 15 (2): 221–240. doi:10.1017/S0047404500000191.

———. 1988. “Methods in the Study of Dialect Perceptions.” *Methods in Dialectology:* 373–395.

Prince, Alan, and Paul Smolensky. 1993. “Optimality Theory: Constraint Interaction in Generative Grammar.” ROA547. http://roa.rutgers.edu/article/view/547.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Reed, David W., and Allan A. Metcalf. 1952. *Linguistic Atlas of the Pacific Coast.* Berkeley: Bancroft Library at the University of California.

Rice, Keren. 1995. “On Vowel Place Features.” *Toronto Working Papers in Linguistics* 14:73–116.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2015. *FAVE (Forced Alignment and Vowel Extraction) Program Suite V1.2.2 [Computer Software].* doi:10.5281/zenodo.22281.

Savage, Matthew, Alexander Mason, Monica Nesbitt, Erin Pevan, and Suzanne Evans Wagner. 2016. “Ignorant and Annoying: Inland Northerners’ Attitudes Toward NCS Short-O.” Poster presented at the American Dialect Society Annual Meeting, Washington, DC.

Scobbie, James M., and Joanne Cleland. 2017. “Area and Radius-Based Mid-Sagittal Measurements of Comparative Velarity.” Paper presented at Ultrafest VIII, Potsdam, Germany.

Scobbie, James M., Eleanor Lawson, and Jane Stuart-Smith. 2012. “Back to Front: A Socially-Stratified Ultrasound Tongue Imaging Study of Scottish English /u/.” *Rivista di Linguistica* 24 (1): 103–148.

Smiljanić, Rajka, and Ann R. Bradlow. 2005. “Production and Perception of Clear Speech in Croatian and English.” *The Journal of the Acoustical Society of America* 118 (3): 1677–1688. doi:10.1121/1.2000788.

Steriade, Donca. 2001. “The Phonology of Perceptibility Effects: The P-Map and Its Consequences for Constraint Organization.” Ms., UCLA.

Stevens, Kenneth N. 1972. “The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data.” In *Human Communication: A Unified View,* edited by Edward E. David and Peter B. Denes, 51–66. New York: McGraw-Hill.

———. 1989. “On the Quantal Nature of Speech.” *Journal of Phonetics* 17:3–45.

Stevens, Kenneth N., Samuel Jay Keyser, and Haruko Kawasaki. 1986. “Toward a Phonetic and Phonological Theory of Redundant Features.” In *Invariance and Variability in Speech Processes,* edited by Joseph S. Perkell and Dennis H. Klatt, 426–449. Psychology Press.

Stone, Maureen, and Eric Vatikiotis-Bateson. 1995. “Trade-Offs in Tongue, Jaw, and Palate Contributions to Speech Production.” *Journal of Phonetics* 23 (1-2): 81–100. doi:10.1016/S0095-4470(95)80034-4.

Sumby, W. H., and Irwin Pollack. 1954. “Visual Contribution to Speech Intelligibility in Noise.” *The Journal of the Acoustical Society of America* 26 (2): 212–215. doi:10.1121/1.1907309.

Sumner, Meghan, and Arthur G. Samuel. 2009. “The Effect of Experience on the Perception and Representation of Dialect Variants.” *Journal of Memory and Language* 60 (4): 487–501.

Thomas, Erik R. 1989. “The Implications of /o/ Fronting in Wilmington, North Carolina.” *American Speech* 64 (4): 327–333. doi:10.2307/455724.

———. 2001. *An Acoustic Analysis of Vowel Variation in New World English.* Publications of the American Dialect Society 85. Durham, NC: Duke University Press.

Thompson, Laurence C. 1987. *A Vietnamese Reference Grammar.* Honolulu: University of Hawaii Press.

Traunmüller, Hartmut, and Niklas Öhrström. 2007a. “Audiovisual Perception of Openness and Lip Rounding in Front Vowels.” *Journal of Phonetics* 35 (2): 244–258. doi:10.1016/j.wocn.2006.03.002.

———. 2007b. “The Effect of Incongruent Visual Cues on the Heard Quality of Front Vowels.” In *Proceedings of the 16th International Congress of Phonetic Sciences,* edited by Jürgen Trouvain and William J. Barry, 721–724. Saarbrücken: Universität des Saarlandes.

Twist, Alina, Adam Baker, Jeff Mielke, and Diana Archangeli. 2007. “Are ‘Covert’ /ɹ/ Allophones Really Indistinguishable?” *University of Pennsylvania Working Papers in Linguistics* 13 (2): 207–216. http://repository.upenn.edu/pwpl/vol13/iss2/16.

Uldall, Elizabeth. 1958. “American ‘Molar’ R and ‘Flapped’ T.” *Bulletin of Laboritorio de Fonetica Experimental da Faculdade de letras da Universidade de Coimbra:* 103–106.

Valkenier, Bea, Jurriaan Y. Duyne, Tjeerd C. Andringa, and Deniz Baskent. 2012. "Audiovisual Perception of Congruent and Incongruent Dutch Front Vowels." *Journal of Speech, Language, and Hearing Research* 55 (6): 1788–1801. doi:10.1044/1092-4388(2012/11-0227).

Wagner, Suzanne Evans, Alexander Mason, Monica Nesbitt, Erin Pevan, and Matt Savage. 2016. "Reversal and Re-Organization of the Northern Cities Shift in Michigan." *University of Pennsylvania Working Papers in Linguistics* 22 (2).

Watkins, Calvert, ed. 2000. *The American Heritage Dictionary of Indo-European Roots.* 2nd ed. Boston, MA: Houghton Mifflin.

Weenink, David. 2014. *Speech Signal Processing with Praat.* http://www.fon.hum.uva.nl/david/sspbook/sspbook.pdf.

Wells, J. C. 1982. *Accents of English.* Cambridge: Cambridge University Press.

Welmers, William E. 1962. "The Phonology of Kpelle." *Journal of African Languages* 1 (1): 69–93.

Winitz, Harris, M. E. Scheib, and James A. Reeds. 1972. "Identification of Stops and Vowels for the Burst Portion of /p, t, k/ Isolated from Conversational Speech." *The Journal of the Acoustical Society of America* 51:1309–1317. doi:10.1121/1.1912976.