UC Berkeley
Dissertations, Department of Linguistics

Title

Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese

Permalink

https://escholarship.org/uc/item/0jr0010h

Author

Faytak, Matthew D

Publication Date

2018-07-01

eScholarship.org Powered by the California Digital Library
University of California

Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese

by

Matthew Donald Faytak

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Keith A. Johnson, Chair
Professor Susan S. Lin
Professor John F. Houde
Professor Richard Ivry

Summer 2018

Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese

Copyright 2018
by
Matthew Donald Faytak

Abstract

Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese

by

Matthew Donald Faytak

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Keith A. Johnson, Chair

This thesis explores the role of uniformity of speech articulation in shaping phonological systems of contrast and their phonetic implementations. The observable effect of uniformity for an individual speaker is that a given phonological primitive (such as a distinctive feature value or gesture, depending on one's theoretical framework) tends to be implemented with maximum articulatory similarity across the speech sounds sharing that primitive. Although less discussed than other organizing principles in substance-based phonology such as phonetic dispersion (Liljencrants and Lindblom, 1972), focalization due to quantal effects (Stevens and Keyser, 1989; Schwartz et al., 1997b), or articulatory ease (Martinet, 1955; Lindblom, 1990), uniformity has been observed in a range of the world's languages, mainly in the timing of laryngeal articulations in stop inventories (Keating, 2003; Chodroff and Wilson, 2017) but also in place-of-articulation primitives (Maddieson, 1996; Chodroff, 2017).

However, uniformity has typically been formulated as a purely linguistic constraint. A primary aim of this dissertation is to motivate uniformity as emerging from domain-general biases that shape complex systems of goal-oriented action more broadly, thereby shedding light on the substantive basis and structure of phonological systems. To this end, I describe a model in which articulatory uniformity emerges from articulatory reuse during learning. During the language acquisition process, a learner's internal model (mapping the effects of motor controls applied to the speech articulators to their outcomes) is not yet fully developed. Under these conditions, a "model-free" learning strategy based on bootstrapping off of the learner's already-mastered skills (exploitation, rather than exploration) may predominate, such that phonological categories whose outputs are perceptually similar may come to be produced with the same articulatory primitives.

This thesis tests aspects of the model of uniformity-through-reuse with an experiment on Sūzhōu Chinese, whose fricative vowels are known to somewhat resemble alveolopalatal fricative consonants in their tongue-palate constriction patterns and fricative noise production targets (Ling, 2009). Ultrasound tongue imaging was used to characterize the typical fricative vowel and alveolopalatal fricative consonant productions of 43 Sūzhōu Chinese speakers.

Analysis reveals that most Sūzhōu Chinese speakers typically use a single tongue posture uniformly across the fricative vowels and consonants examined, while a minority of speakers deviate from uniformity to an idiosyncratic extent. The extent to which a speaker deviates from a uniform strategy is shown to be unrelated to demographic characteristics and language ability in Sūzhōu Chinese and Standard Chinese. This pattern at the population level suggests that "speaker-side" factors, such as articulatory reuse, are primarily responsible for shaping the "synchronic pool of variation" (Ohala, 1989) for this set of Sūzhōu Chinese segments.

To Natalia

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation would not have been possible without the strong support network that surrounded me during my time as a graduate student, both at Berkeley and beyond, and particularly after a head injury in 2017 put me into a cognitive fog for more than a year in the middle of the dissertation-writing process. I cannot possibly hope to fully represent this staggering generosity in two or three pages. But try I must.

I first thank the academic advisors I have worked with the most closely, Keith Johnson and Susan Lin. I thank Keith for his supportive presence and his exceptional willingness to answer a knock at the door at seemingly any time; I thank Susan for introducing me to ultrasound tongue imaging and for always pushing me towards greater precision in my experimental noodling. Both have been generous with their technical prowess and practical advice. I can only hope to take after their advising and work to point my small corner of the field in a more supportive, mindful, and equitable direction.

I also extend my thanks to Ron Sprouse, who along with my advisors-on-paper was instrumental in developing and directing my computational skills before and during the writing of this thesis. While he is already noted around the department for his general technical wizardry, he is also an immensely competent teacher of Python, which I can only hope to emulate going forward.

My other committee members, John Houde and Rich Ivry, deserve recognition for their timely and incisive feedback; I am especially grateful to John for our lengthy discussions on speech motor control, which I hope to keep up in the future.

I am grateful to Sharon Inkelas for convening the Social Science Matrix working group on similarity back in 2014, which in many ways sparked my interest in the topic explored in this dissertation. I owe Larry Hyman thanks for much interesting discussion on phonology and Grassfields Bantu, and besides this, a great deal of wine, food, and hospitality ("If he didn't exist, we'd have to invent him!").

I also must thank the individuals who did the most to create and maintain the convivial atmosphere in the Linguistics Department. Among members of the PhonLab, I thank Auburn Barron-Lutzross, Sarah Bakst, Andrew Cheng, Jevon Heath, Emily Remirez, and Alice Shen for being the best office-mates one could hope for. Down on the A level, other department members did much to enliven the sometimes monastic grad school experience: Nico Baier, Kenny Baclawski, Jack Merrill, Kelsey Neely, Nik Rolle, Katie Sardinha, and Elise Stickles, among others.

Since the department would most likely plunge into chaos and darkness without their work, I am especially indebted to Belén Flores and Paula Floro, who have helped me and countless others navigate administrative hurdles large and small, particularly after my head injury.

Beyond Cal, other colleagues have contributed to the development of the work presented here in a number of ways. I am particularly grateful to Eleanor Chodroff for our discussions of the overlap between our dissertations, which has greatly clarified the presentation of ideas here. Jeff Mielke merits special mention for technical advice and inspiration. I also owe much

to the sponsorship and assistance of the faculty and students at Fudan University. I must especially thank Chen Zhongmin for taking the time to sponsor my travel, and Wang Feifan for his logistical help and invaluable advice on doing linguistic research in his hometown.

Finally, I thank my mother, Coleen, for making all of this possible: your confidence in me has always been among my best resources. And I thank my wife, Natalia, for joining me on travels already taken and yet to come: amid all the changes, your companionship has been a stabilizing constant and a truly valuable thing to me.

Chapter 1

Introduction

The phonological inventories of the world's languages are characterized by regular structure. A great deal of this structure, at various levels of representation, can be attributed to major principles discovered through work in phonetics and substance-based phonology: articulatory ease (Martinet, 1955; Lindblom, 1990), perceptual dispersion (Liljencrants and Lindblom, 1972; Lindblom, 1990; Flemming, 2004), and focalization due to quantal effects (Stevens and Keyser, 1989; Schwartz et al., 1997b). Some recurring structure in phonological inventories cannot, however, be modeled using these factors alone: none of these factors explain why phonological inventories favor internal regularity in their language-specific phonetic implementations of phonological content (Ohala, 1979; Schwartz et al., 2007). Only an additional principle can explain some of this regularity. This dissertation concerns itself with this additional principle, which I dub UNIFORMITY, a name reflecting its observable consequences on linguistic systems that is also shared by a small range of papers on the topic (Keating, 2003; Chodroff, 2017; Chodroff and Wilson, 2017).

The goal of this dissertation is to assemble a comprehensive account of uniformity's contribution to regular phonological structure and its implications for researchers in phonetics, as well as to propose a model of how it emerges. Other substantive pressures that exert their effects on phonological structure can be taken as emerging from the physical circumstances of speech production and perception (Lindblom, 1984; Schwartz et al., 1997a; Schwartz et al., 2007). But whence uniformity, which often appears to exist purely for uniformity's sake and often counteracts the effects of other inventory-shaping substantive pressures like dispersion? I attribute uniformity, at least in terms of the articulatory outputs of the speech planning and execution process, to articulatory reuse, or the tendency to learn new task-directed actions by repurposing components of related actions that are known to lead to success.

While directly testing for the occurrence of articulatory reuse during learning is beyond the scope of this dissertation, the reuse model for uniformity does make some principled predictions about which types of phonological subsystems should tend to exhibit articulatory uniformity as a consequence of articulatory reuse in adult phonological systems. The experimental component of this dissertation examines one such phonological subsystem, the set of FRICATIVE VOWELS and fricative consonants in Sūzhōu Chinese, for uniformity. The resulting production experiment, which directly assesses uniformity of articulatory strategies using ultrasound imaging, aims to add to the case studies of uniformity already in existence with its focus on direct articulatory assessment and its exploration of an otherwise unexplored type of phonological subsystem.

1.1 Main ideas

This thesis advances two theoretical concepts based on review of existing literatures in linguistics (phonetics and phonology) and motor control. One concept, that uniformity constrains variation of phonetic outputs relative to one another, is an analytical point I share with existing literature on uniformity and "structured variation;" the other, that uniformity is a consequence of reuse during learning, is articulated here for the first time, to my knowledge. I discuss each concept in turn below.

1.1.1 Uniformity constrains phonetic implementation of phonological material

I take as a starting point for the discussion below that there is some division between phonological material (mental representation of action) and phonetic implementation (action itself), counter to Ohala (1990) but in line with other, more modular accounts of phonetics and phonology (Keating, 1984; Keating, 1990). Phonological material, at the level of the feature or gesture, is constrained in its mapping into phonetic substance: outputs that share aspects of their mental representations, in terms of phonological primitives, are permitted to differ only to a limited extent in some articulatory or acoustic output parameters (Keating, 1990). This restriction on inter-category variability in phonetic implementation, and the resulting covariation of sounds sharing some phonological material in aspects of their phonetic realization, I dub UNIFORMITY in line with a small but growing line of research on the topic (Keating, 2003; Chodroff, 2017; Chodroff and Wilson, 2017).

Generally, uniformity is modeled as a specifically linguistic organizational principle (Keating, 2003; Chodroff, 2017) that is amenable to explanation in a constraint-based framework, given that it is a violable tendency that speakers appear able to prioritize over other organizational principles that may affect phonetic outputs (Keating, 2003). There is generally no further exploration of the origins of this principle, however, and uniformity as a force exerted on complex systems of action is typically formulated as specifically linguistic. In keeping with the general mission of substance-based phonology to "derive language from non-language" (Lindblom, 1984), and unlike existing accounts, I propose a behavioral origin for uniformity in how language learners build on successful articulatory strategies through reuse during language acquisition.

1.1.2 Uniformity arises from articulatory reuse during learning

Locating the aspect of "non-language" that is responsible for the observed patterns in phonetic outputs leads me to explore of the motor control literature in a search for potential behavioral underpinnings of uniformity. I localize this aspect of behavior to language development, and more specifically in how speech motor learning operates during language development. In short, developing L1 speakers reuse aspects of their limited articulatory repertoire to expand the number of contrastive segments they can readily and consistently produce (so as to more reliably have communicative success).

I put forward that this ARTICULATORY REUSE comes about due to limitations on the learner's ability to produce adult-like speech. In part, and more obviously, this is due to limited motor skill on the part of child language learners. Less obviously but importantly, language learners also lack a well-developed acoustics-articulation-motor mapping, or INTERNAL MODEL, which could be used to adjust the controls applied to the speech production plant (the vocal tract) based on predictions of the sensory consequences of motor commands (Wolpert et al., 1995; Wolpert and Flanagan, 2001; Houde and Nagarajan, 2011). In place of this implicit, model-driven type of learning, which has been thoroughly researched in adult speakers (Houde and Jordan, 1998; Shadmehr and Krakauer, 2008), learners may instead default to a process of explicit (conscious) selection of articulatory strategies, which is known to be flexibly combined with model-based learning processes (Haith and Krakauer, 2013; McDougle et al., 2016) but may predominate when estimates of the consequences of actions are unreliable (Loeb, 2012).

Conscious selection of articulatory strategy opens the door to a particular solution to the EXPLORE-EXPLOIT DILEMMA inherent to complex decision-making: speakers can either continue employing mastered strategies to achieve a task, even if they are suboptimal, or strike out and explore the task space in search of a global optimum (Wilson et al., 2014). Articulatory reuse, then, is the decision to continue *exploiting* existing articulatory routines that are known to result in success, rather than engage in a thorough search of the task space for global optima. Exploiting one control for achieving two tasks is simply more efficient, in terms of energy expended in the search process, than using two controls for two tasks if they have similar goals, such that speakers can successfully use one control rather than multiple controls. The phenomenon of articulatory reuse is thus grounded in effort reduction (Zipf, 1949), which is far from specifically linguistic.

If this exploitation crosses over into exploitation of a single learned routine for multiple segments with similar or related goals, such as vowels of a given height (Ménard et al., 2008) or stops with similar voice-onset times (Chodroff and Wilson, 2017), then the activation basis of these sets of sounds may become inextricably linked. Although I do not directly study child learners in this dissertation, I posit that one consequence of articulatory reuse during learning is articulatory uniformity in the adult language, provided that the use of shared articulatory primitives discussed above persists through development. So, uniformity, if it does derive from articulatory reuse, is not specifically linguistic in origin, but rather emerges from a non-linguistic bias toward effort reduction during motor learning.

1.2 Aims of the thesis

The two aims of this thesis are as follows: first, to provide a comprehensive review of uniformity-like principles in linguistic science, and (to my knowledge) the only review of emergent uniformity-like principles across linguistics and motor control (speech and otherwise). Secondly, I aim to experimentally verify some predictions of my model of uniformity-via-reuse. This is accomplished through a production experiment on adult speakers of Sūzhōu Chinese, in which uniformity across a set of sounds that appear to share fricative noise targets is assessed directly from ultrasound video data.

1.2.1 Consolidating evidence for uniformity

A major goal of this dissertation is to assemble an account of uniformity's behavioral effects and its origins. Unlike previous accounts of uniformity, here, I reach beyond strictly linguistic evidence to incorporate evidence from research in motor control on the learning, consolidation, and updating of complex, task-directed behaviors in general. In a sense, the model outlined in Chapter 2 could be taken as a general overview of emergent structure in large, interrelated sets of task-directed actions, but is primarily applied to speech motor learning and accordingly targeted at an audience mostly consisting of linguists.

I draw linguistic evidence for uniformity from several threads of research. A review of the expanding literature on uniformity and the related concept of "structured variation" in phonetics (Keating, 2003; Chodroff, 2017; Chodroff and Wilson, 2017) confirms the basic validity of the pattern of uniformity. I compare this formulation of the organizational principle to some forebears in substance-based phonology, including GESTURE ECONOMY as articulated in (Maddieson, 1996) and several models of the emergence of phonological structure, primarily articulated by Lindblom and colleagues (Lindblom, 1998; Lindblom et al., 2011) but also including others (Ohala, 1979; Schwartz et al., 2007). Evidence from child language research of articulatory reuse within limited speech production repertoires is also considered (Menn, 1978; Menn, 1983; Vihman, 2014). In particular, I focus on the need for learners to be faithful to adult targets in spite of their limited articulatory capabilities and underdeveloped internal models, as articulated in McAllister Byun et al. (2016).

I additionally turn to speech motor control to clarify aspects of the internal models described above, their role in different types of learning, and the contributions of different styles of learning to the overall speech motor learning curve. I outline the division between model-free (or cognitive) learning that adjusts motor controls rapidly on the basis of explicit selection of actions, on the one hand (Haith and Krakauer, 2013; McDougle et al., 2016; McDougle et al., 2015), and model-driven learning that primarily relies on an internal model to update the controller below the level of consciousness, on the other hand (Houde and Jordan, 1998; Todorov and Jordan, 2002; Houde and Nagarajan, 2011). Model-free learning is shown to produce behaviors consistent with articulatory reuse, and is shown to be likely to occur under conditions consistent with those encountered by the learner during language acquistion.

1.2.2 Experimental verification of aspects of uniformity

According to the model briefly articulated in Section 1.1.2, speech sounds that sound alike to the learner but are accomplished by the adult model using distinct articulatory strategies will tend to be "mis-acquired" as using an identical articulatory primitive to some other sound or series of sounds. Sūzhōu Chinese presents a case in which this type of acquisition may have a particularly large impact on an easily observable aspect of articulation, tongue posture. It stands to reason that if speech production strategies are developed through articulatory reuse, they remain (in some sense) yoked to one another throughout development and into adulthood.

Sūzhōu Chinese has in its inventory, among other speech sounds, a series of fricative consonants at an alveolopalatal place of articulation (e.g., /ɕ/) and two FRICATIVE VOWELS (one rounded and one unrounded) that exhibit fricative noise resembling that produced by the /ɕ/ series of consonants. On the basis of prior research and pilot data, the fricative vowels appear to have two major articulatory variants that do not have reliable acoustic correlates. An apparently conservative variant exhibited by older speakers could be described as having a dorso-postalveolar tongue-palate constriction pattern (Ling, 2009), akin to a hyperarticulated [i] but quite distinct from that vowel. An apparently innovative variant instead exhibits a lamino-postalveolar constriction of undetermined type (Ling, 2009). It is an open empirical question whether this second variant is uniform in lingual articulatory strategy with the /ɕ/ series of consonants, but prior research strongly suggests that this is likely (Ling, 2007; Ling, 2009). I speculate that this uniformity could be due to reuse of the tongue position from the /ɕ/ series of consonants for fricative vowel production, on the basis of a common acoustic goal, [ɕ]-like fricative noise.

The results of a production experiment to characterize the lingual articulation of these phones for 44 speakers suggest that most speakers favor a uniform lingual posture for the fricative consonants and vowels as a set. There is plenty of individual articulatory idiosyncrasy, however: many speakers exhibit a relatively /i/-like lingual posture during articulation of the fricative vowels that is nonetheless distinct from both the /ɕ/ series of fricatives and the high front vowels. This variation has an unclear correspondence to acoustic outputs: so long as the constriction location does not substantially change, there are no major effects on the fricative noise produced, and there are extremely small changes to the formant frequencies observed for the vowel.

The experimental results also show that uniform strategies for fricative vowel production do not come to predominate in the Sūzhōu Chinese-speaking population over time, which initially appears to be a strike against the theorized role of reuse in producing uniform strategies. Learners should tend to favor uniform strategies in acquiring series of speech sounds; if adult models (who themselves favored uniform strategies during language acquisition) present learners with a uniform articulatory strategy for this series, and this strategy has overt acoustic consequences, then the speaker population should gradually come to favor uniformity over time. However, the absence of change over time may simply be due to variation in the Sūzhōu Chinese fricative vowels' articulatory strategies being largely covert, which restricts the impacts of uniformity to the individual learner.

Regardless of the specific outcome of this study, an additional benefit of the Sūzhōu Chinese test case is to expand the empirical reach of observations relating to a theory of uniformity. Research on uniformity is, at the present moment, somewhat lacking in empirical breadth, with most case studies and associated data sets discussing English stop consonants (Flege, 1982; Keating, 2003; Chodroff and Wilson, 2017). Only limited exploration of other languages was found in the literature reviewed (Ewe in Maddieson, 1996; Czech in Chodroff, 2017). The Sūzhōu Chinese case study the empirical breadth of observations of articulatory uniformity, not only by examining the phonetics of an understudied language, but also through the focus on the unusual class of segment discussed above, the fricative vowel.

1.3 Outline

This thesis is structured as follows. This chapter provides an overview and outline. Chapter 2 reviews the evidence for linguistic structure not attributable to principles already treated in substance-based phonology, and ultimately gives an account of articulatory uniformity as emerging from articulatory reuse. Principles discussed and disambiguated in Chapter 2 include notions such as uniformity, economy, and symmetry. Chapter 2 also introduces the motor control literature in some detail, which makes it plain that the learning of complex systems of task-directed action (such as speech) may introduce structural regularities into the system that results via trial-and-error learning and explicit formulation of articulatory strategy by the learner. This component of the learning process is connected to the child language literature before the chapter concludes.

Chapters 3 and 4, taken together, introduce the Sūzhōu Chinese case study. Chapter 3 provides background on Suzhou Chinese and prior research on its phonetics and phonology, focusing on attributes of the language that are relevant for experimental design. Some recent historical background on the development of the language, based on a brief historical-comparative analysis, is also provided. Chapter 3 takes as a secondary aim the expansion of the vowel space beyond its commonly assumed boundaries to include vowels with non-dorso-palatal constrictions. This situates the Sūzhōu Chinese fricative vowels, which are produced with postalveolar constrictions made either by the tongue blade or body, as vowels proper, rather than some other type of segment, although the analyses in Chapter 4 do not crucially depend on this analytical point.

Chapter 4 describes the Sūzhōu Chinese production experiment itself. First, a pilot study carried out at the University of California, Berkeley on a small number of Sūzhōu Chinese speakers is discussed, given its influence on the hypotheses that are investigated in the full-length experiment. After discussion of hypotheses, the study methods are described, with a particular focus on the linguistic materials. Acoustic and articulatory results (from ultrasound imaging) are each described in turn, with a focus on the latter. The articulatory results are obtained primarily through dimensionality reduction methods applied directly to preprocessed ultrasound frames; automatic classification algorithms are then applied to the data to assess similarity of the fricative vowels to the fricative consonants in a principled, data-driven way. The findings are discussed in terms of their implications for the uniformity and reuse framework developed in previous chapters. Some attention is given to the fact that uniformity in Sūzhōu Chinese is largely covert in nature, with little to no impact on the portions of the acoustic signal that speakers seem likely to attend to.

Chapter 5 summarizes the findings of the dissertation and gives directions for future research, including several possible experimental methods for confirming aspects of the reuse-and-uniformity framework not directly tested in the Suzhou Chinese experiment. These include direct assessment of the learning curve in speakers learning a novel speech sound on the one hand and longitudinal studies of learners on the other hand. A parallel between L1 learners and L2 learners is explicitly considered: it stands to reason that similar pressures toward articulatory reuse apply in both situations, and could cause articulatory uniformity within an L1 to resemble L1-L2 uniformity in some senses. Some implications of uniformity are discussed, including the possibility of its impacts on sound change, in spite of the fact that its direct impact on articulatory strategy may often be covert. Impacts on higher-level phonological structure, which may emerge over generations of language transmission subject to reuse during learning, are also considered.

Chapter 2

Articulatory uniformity and articulatory reuse

This chapter lays out the theoretical background and models behind the core hypothesis explored in this dissertation: that speech motor programs are constrained in their variability with respect to one another if they have similar acoustic goals. Throughout, I will refer to this constrained variation as ARTICULATORY UNIFORMITY, borrowing the latter term from prior research on covariation in articulatory strategies (Maddieson, 1996; Lindblom, 1998; Lindblom et al., 2011) and acoustic targets (Chodroff and Wilson, 2017; Chodroff, 2017). Attested examples of articulatory uniformity involve a series of sounds with (a subset of) tightly covarying articulatory strategies consistent within an individual speaker, such as sets of front and back vowels of the same tongue height (Ménard et al., 2008), stops at several places of articulation with the same voice-onset time (Chodroff, 2017; Chodroff and Wilson, 2017), and fricatives with the same spectral prominences across voicing types (Chodroff, 2017). This is not to be confused with consistency of outputs across speakers; the same studies that comment on within-speaker uniformity also often comment on the inter-speaker variability in terms of uniform output targets (Ménard et al., 2008; Chodroff, 2017).

I refer to the apparent cause of uniformity as ARTICULATORY REUSE. In the absence of a reliable INTERNAL MODEL of how motor commands, articulation, and the resulting acoustic outputs are related, learners may instead turn to directed exploration of their established articulatory routines, resulting in a range of successful strategies that simply "hack" and reuse articulatory strategies already used elsewhere (Loeb, 2012; McAllister Byun et al., 2016). Learners are known to explore within the bounds of their established articulatory routines by trial and error to achieve goals (Wilson et al., 2014); this is particularly true during language acquisition (McAllister Byun et al., 2016). In motor learning in general, learners' performance is also known to be better within a task space they have already mastered (Braun et al., 2009; Ranganathan et al., 2014) or using movements they habitually carry out (Diedrichsen et al., 2010). I accordingly argue that a bias towards articulatory reuse during trial-and-error learning of speech motor programs results in task variables for speech production that remain "yoked" across series of speech sounds into adulthood. Under this approach, the driving forces behind uniformity are not specifically linguistic (and thus part of the language faculty to the exclusion of other neuroanatomical or behavioral modules), but are rather a consequence of the neurological and behavioral bases of learning more generally.

In this chapter, I elaborate a model of articulatory reuse and articulatory uniformity and provide justification for connecting these phenomena. In Section 2.1, I provide background on the concept of uniformity as it has sporadically arisen in research on speech production: while rarely formulated in exactly the same way in any two different programs of research, some common observations on the phenomenon can be gleaned. Beginning with Section 2.2, I articulate my own framework of uniformity as arising from articulatory reuse, primarily as a side effect of the learning of speech motor programs during language acquisition. I consider the phonological primitives that may be the basis of reuse in Section 2.2, the biases associated with learning that motivate articulatory reuse in Section 2.3, and the critical junctures at which these biases are expected to dominate learning outcomes in Section 2.4. Section 2.5 synthesizes and summarizes the content and establishes the usefulness and relevance of the particular experiment undertaken in Chapter 4.

2.1 Background: economy and uniformity principles

Substance-based phonology has long aimed to explain why individual speakers' repertoires of vowel and consonant articulatory strategies utilize a relatively small number of acoustic dimensions to produce a relatively large number of contrasts, even when this results in a less well-dispersed inventory than might otherwise be possible (Ohala, 1979; Schwartz et al., 1997a; Schwartz et al., 2007; Lindblom et al., 2011). In this section, I provide an overview of various proposed formulations of regularity in phonological and phonetic structure. First, I present an overview of dispersion as an explanatory factor in phonological structure, as well as its limitations, in Section 2.1.1. I then present economy principles (Section 2.1.2), which take the efficient recombination of phonological primitives as their focus, contrasting them with the main target for discussion, uniformity principles in their various proposed forms (Section 2.1.3). Some shortcomings of both principles are discussed in Section 2.1.4.

2.1.1 Dispersion theory and its limitations

Phonological inventories generally exhibit a great deal of regular structure, at least some of which appears to come at the cost of optimized perceptual dispersion of an inventory's constituent phonemes in perceptual space. Dispersion Theory, which predicts that languages will maximize (or at least make sufficient) acoustic separation of phonological categories, is among the earliest contributions made by substance-based phonology to the study of language structure (Liljencrants and Lindblom, 1972). Subsequent developments of the theory incorporate findings associated with Quantal Theory (Stevens and Keyser, 1989), which predict that some vowels that exhibit focalization of two formants (such as [i], [a], [u]) will be more commonly attested in vowel inventories (Schwartz et al., 1997b). Computational

	Front		Central		Back	
High	i	ĩ	ɯ	ɯ̃	u	ũ
Mid-high	e		ə		o	
Mid-low	ɛ	ɛ̃	ʌ	ʌ̃	ɔ	ɔ̃
Low			a	ã		

Table 2.1: Monophthongal vowel phonemes of Acehnese (Asyik, 1982).

	Front		Central		Back	
High	i	ĩ	ɯ	ɯ̃	u	ũ
Mid-high	e		ɤ		o	
Mid-low	ɛ	ɛ̃	ʌ	ʌ̃	ɔ	ɔ̃
Low	æ	æ̃	a	ã	ɑ	ɑ̃

Table 2.2: Vowel phonemes of Dan Blossé (Vydrine, 2004).

simulation of vowel systems using some combination of dispersion and focalization effects generally reproduces a substantial portion of the typology of vowel systems (Schwartz et al., 1997b; Schwartz et al., 1997a; Berrah and Laboissière, 1997; de Boer, 2000).

However, a long-standing observation of functionally oriented linguistics has been that sounds exist in series, which share phonetic or phonological characteristics among themselves (see Martinet, 1955), rather than as uncorrelated arrangements of highly distinct categories that maximize dispersion (Ohala, 1979; Schwartz et al., 1997b). Dispersion and focalization alone in fact predict different, maximally dispersed vowel systems, which do not capture much of the existing vowel system typology; in particular, it is only when a system has more than nine vowels, most frequently a set /i, ɪ, e, ɛ, a, ɔ, o, ʊ, u/ or similar, that secondary sets of nasalized or differently phonated vowels typically develop (Schwartz et al., 2007). Languages instead tend to utilize a handful of acoustic dimensions with great efficiency to implement phonological category contrasts. In some respects, even the crowded vowel inventories of languages such as Acehnese (Table 2.1) and Dan Blossé (Table 2.2) are unsurprising, given their orderly use of a handful of acoustic dimensions for vowel contrasts: all central and front vowels are unrounded and nasality is combined systematically with nearly all vowel qualities.[1]

These very densely populated but highly structured systems stand in contrast to the reasonably well-dispersed but structurally lacking systems of Hopi (Table 2.3) and Nivkh (Table 2.4), which are quite typologically unusual. Hopi exhibits two apparently distinct

---

[1] Even the absence of mid-high nasal vowels in both languages is consistent across backness-rounding series and has a perceptual explanation: nasality intrinsic to vowels tends to lower the perceived height of the vowel (Krakow et al., 1988), such that the contrast with the mid-low nasalized vowels is more difficult to maintain.

	Front		Central		Back	
	Unrd.	Rd.	Unrd.	Rd.	Unrd.	Rd.
High	i		ɨ			
Mid		ø				o
	ɛ					
Low			a			

Table 2.3: Vowel phonemes of Hopi (Jeanne, 1978).

	Front		Central		Back	
	Unrd.	Rd.	Unrd.	Rd.	Unrd.	Rd.
High	ɪ					u
Mid	i͡e				ɤ	o
Low	æ					

Table 2.4: Vowel phonemes of Nivkh (Maddieson and Disner, 1984).

mid heights for front rounded and unrounded vowels, and only one back vowel among its six. In Nivkh, which Maddieson and Disner (1984) describe as “defective” in its use of the acoustic space available for vowel production, only one vowel is diphthongized, a rounded-unrounded pair at a given height and backness occurs only for the back vowels, and the heights occupied by high and mid vowels in the front and the back of the vowel space differ. In both the Nivkh and Hopi vowel inventories, it is harder to observe any recurring surface similarities among phonemes, like the fully rounded-out combinations of height, backness, and nasality contrasts observed in Acehnese and Dan Blossé. That vowel systems like this (highly dispersed and featurally complex) are rare suggests that perceptual dispersion is only one of a set of factors that determines the phonological structure of vowel systems.

Beyond phonological inventory structure, there are also some aspects of phonetic structure that are not attributable to dispersion. For example, in Keating (2003)’s discussion of interacting linguistic and non-linguistic influences on speech production, she includes a reanalysis of data from Flege (1982) on the variable production of prevoicing in utterance-initial American English voiced stops /b, d, g/, which are known to be variably produced as short-lag stops or with prevoicing (see also Keating, 1984; Docherty, 1992; Davidson, 2016). Flege (1982) observes that the articulatory strategies for stop prevoicing employed by individual speakers ($n = 10$) exhibit inter-speaker variation in the timing of vocal fold adduction that is largely idiosyncratic. He makes another important observation: a majority of the participants engage in adduction of the vocal folds before or during stop closure regardless of whether it has any acoustic consequences, namely the initiation of prevoicing during the stop closure. This is reminiscent of other findings that laryngeal activity in stops is often “stereotyped” or remarkably insensitive to context (Löfqvist and Yoshioka, 1981).

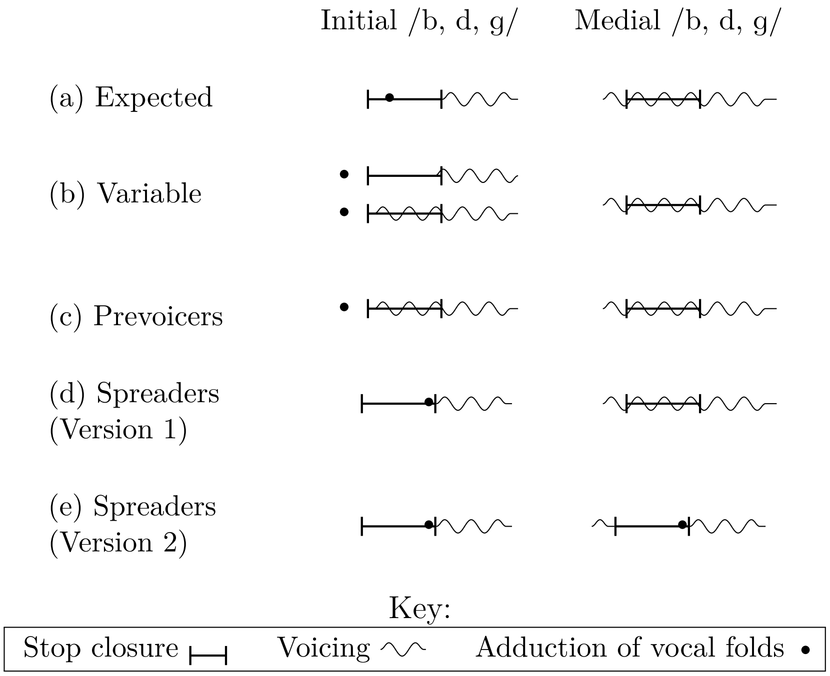


Figure 2.1: Attested individual articulatory strategies for short-lag VOT ("voiced") stops in English. Data from Flege (1982), as analyzed by Keating (2003).

Keating (2003) identifies four groups among Flege (1982)'s ten subjects, which I schematize in Figure 2.1 using a modified version of Flege (1982)'s Figure 4 with an additional schema for the medial context that Keating brings up by way of comparison. While the number of speakers in each group is small, some qualitative judgments can readily be drawn from the data. One speaker, who Keating classifies as a group unto themselves, patterns as Keating expects (Figure 2.1a). Adduction of the vocal folds (as measured with an electroglottograph) occurs around the start of the stop closure in initial (phonologically) voiced stops. Keating attributes this to the absence of other articulatory maneuvers required for prevoicing in this position (376). This speaker thus produces two positional variants of each voiced stop: one fully prevoiced, the other with short-lag voicing, with different associated timing of vocal fold adduction.

The remaining three groups' behaviors are less expected. One group ($n = 3$) varies in terms of when voicing is initiated, but adduction of the vocal folds always occurs before the stop closure, regardless of the timing of voicing (Figure 2.1b). Flege (1982) describes the variation as essentially bimodal. This pattern is more uniform in terms of articulation across the positional variants—both initial and medial /b, d, g/ share full adduction of the vocal folds—but requires more effort in initial position and does not result in uniform acoustic output.

The other groups are even more consistent across their positional variants. Keating observes that another group, dubbed the Prevoicers ($n = 4$), ensures acoustic and articulatory similarity of initial and medial voiced stops by consistently adducting early and making all articulatory adjustments required for full prevoicing, a simple but especially effortful strategy that results in full vocal fold adduction and full prevoicing in all contexts (Figure 2.1c).

The final group ($n = 2$), dubbed the Spreaders, counterintuitively exhibits spread vocal folds during onset voiced stop closure, adducting just in time to avoid producing longer voicing lag that might give the percept of an aspirated stop. Assuming that medial stops are voiced for these speakers, the Spreaders exhibit two very distinct positional variants for their voiced stops (Figure 2.1d). On the other hand, although 80–90% of medial /b, d, g/ in English are voiced (Flege and Brown, 1982; Keating, 1984; Docherty, 1992; Davidson, 2016), Keating notes (2003, p. 377) that these speakers' medial /b, d, g/ could constitute the small proportion of medial /b, d, g/ that are reported without voicing during the stop closure. In this situation, the Spreaders would exhibit an unusual but more articulatorily and acoustically consistent set of positional variants, in which intervocalic /b, d, g/ are implemented as short-lag stops like the onsets (Figure 2.1e).

In sum, of the ten speakers who participated in Flege (1982), seven effectively adduct their vocal folds during all voiced stops, including in initial position, regardless of acoustic effect. Keating identifies a qualitative difference between the six Prevoicers and Spreaders (supposing medial stop production as in Figure 2.1d-e) and the other speakers. These speakers can be said to "[prefer] uniformity over articulatory ease" (377) in that either initial or medial stops are produced in a way that is contextually effortful but maintains similarity with the other contextual variant. Moreover, the Variable Prevoicers are articulatorily uniform in that they consistently adduct the vocal folds regardless of whether or not voicing is initiated or fails during stop closure. They can thus be said to favor "uniformity of glottal articulation," but not uniformity in acoustic prevoicing (377).

Subphonemic patterns of within-speaker regularity, such as this, are not uncommon. They have also been found for place of articulation for coronals in French and English (Dart, 1998), vowel height (European and Québecois French: Ménard et al., 2008; English: Johnson et al., 1993), and /ɹ/ production in American English (Mielke et al., 2016). This is not intended to serve as an exhaustive list, but rather to underscore that articulatory uniformity is an apparently pervasive fact about language production that is often only noticeable with direct articulatory imaging or articulatory modeling, in the case of Ménard et al. (2008).

2.1.2 Economy principles

Because of the anti-dispersion effects discussed in Section 2.1.1, an additional factor beyond perceptual dispersion appears to be required to constrain the vowel typology to its observed tendency to elaborate contrast along a small number of acoustic dimensions. Economy principles are one such attempt to capture this additional factor, formalized as a tendency to utilize with maximal efficiency the contrastive phonological features in an inventory. An early and influential formulation of economy of phonological features is the principle of

(a)	/p/, /t/, /k/, /b/, /d/, /g/
(b)	/m/, /ts/, /ɗ/, /k’/, /r/, /ɬ/

Figure 2.2: Example consonant inventories congruent with MUAF (a) and not congruent with MUAF (b). Inventory (b) from Ohala (1979).

Maximum Utilization of Available distinctive Features (MUAF), succinctly laid out in Ohala (1979).

MUAF can be summarized with a pair of observations: variants on the consonant system shown in Figure 2.2 (a) are extremely common among the world's languages, even though most of the consonants differ from some other consonant "by a minimum, not a maximum number of distinctive features" (3). On the other hand, the system illustrated in Figure 2.2 (b), in which the consonants are more dispersed in featural terms, would be shocking if discovered as the consonant inventory of a heretofore undocumented language. This inventory breaks with the typological tendency to use a given feature (if it has contrastive value in a phonological system), in combination with other available features to the extent possible.

The foundation for much of this work can be found in André Martinet's *Économie des changements phonétiques* (Martinet, 1955). Observing that the tension between a minimization of effort (Zipf, 1949) and the need to communicate essentially drives language change, Martinet posits contrastive feature economy, the maximal use of as few distinctive features as possible to define as many contrastive speech sounds as possible, as emergent from these two demands (Martinet, 1955, pp. 71–74). Elsewhere, Martinet notes some apparent benefits of economy in terms of ease of production and action programming, as well as learnability:

> [F]or the same total of phonemes, they require less [sic] articulations to keep distinct; these articulations, being less numerous, will be the more distinct; each of them being more frequent in speech, speakers will have more occasions to perceive and produce them, and they will get anchored sooner in the speech of children. (Martinet, 1968, p. 483)

This functional basis of his account is an important contribution to substance-based linguistics, and the need to provide uniformity and economy principles with a basis is a reason that articulatory reuse is linked to articulatory uniformity here.

One attempt to elevate economy to the level of a linguistic desideratum, rather than leaving it as a side effect of language change, is Maddieson (1996)'s proposal for Gesture Economy. Maddieson (1996) examines stop consonant constriction production in Ewe, a language which has in its phonemic inventory simple bilabial and velar stops as well as doubly-articulated labiovelar stops. Using electromagnetic articulography, he determines that speakers of Ewe, in the aggregate, use "very comparable time courses … [trajectory] shapes … and amplitudes" in the movement trajectories of the lips and tongue in producing the two stop closures in labiovelar stops (i.e. [kp], [gb]) and in the single closures of bilabial or velar stops (i.e., [k], [b]) (575–76). This near-exact similarity of movement trajectories need not be the case: speakers might be expected to exhibit some sensitivity to the antagonistic demands of other nearby segments or the simultaneously articulated components of complex segments.

Maddieson (1996) concludes that the Ewe labiovelar stop data point to a principle of Gesture Economy: languages exhibit a tendency to maximally re-use gestures in a maximally large set of the contrastive sounds of the language (p. 574). The "gesture" primitives over which his proposed economy principle operates are explicitly defined: they are not the same as the primitives that happen to share the name "gesture" in the Articulatory Phonology framework (Browman and Goldstein, 1992). Rather, Maddieson (1996) defines these subsegmental units as "typical movement trajector[ies] for a given articulatory subsystem in realizing a given phonetic contrast," controlling for segmental and supersegmental context (p. 574); one could identify the units compared in the analysis at hand as corresponding to place of articulation.

Clements (2003)'s FEATURE ECONOMY is likely the best-known economy principle formalized as a demand on linguistic systems, as well as the most thoroughly tested. Economy is quantified in this proposal as the ratio of the number of phonemes to the number of features needed to describe all oppositions among them. Clements finds statistical evidence in a large set of phonological inventories for two predictions of his model of economy: that series of sounds sharing the implementation of some feature will be favored (2003, p. 296), and that singular implementations of some feature combination ("isolated sounds") will be disfavored (2003, p. 306). Both findings are important confirmations of Martinet's (1968, p. 483) observed tendencies in language change.

Clements (2003)'s formulation of Feature Economy differs in an important way from Maddieson (1996)'s Gesture Economy: the units that languages economically deploy are abstracted features partially divorced from articulatory substance, not articulatorily-grounded gestures. Clements argues that feature economy and gesture economy make different predictions about the typology of phonological inventories: gestures are "overly specific" and do not generate "the full range of economy effects observed in sound systems" (2003, p. 325). He comments explicitly on gesture economy's connection to learnability (à la Martinet), noting that he finds it doubtful that gesture economy would actually aid learnability:[2]

> We have already seen that feature economy cannot be reduced to simple gesture economy; for example, voiceless fricatives show strong economy effects, but the gesture required to guarantee continuous noisy airflow in [coronals] is quite different from that required in [labiodentals] or [velars], and it is difficult to see how the mastery of one could facilitate the acquisition of another at the purely motor level. Nor does the generalisation of continuance from [coronal fricatives] to [labiodental fricatives] and [velar fricatives] increase the token frequency of

---

[2] This notion will be revisited in Section 2.2.5 and 2.3, where I will argue against it: reuse of speech motor programs across multiple speech sounds should make speech motor programming slightly less taxing for adults with native-like control and considerably less taxing for L1 and L2 learners.

> any gesture, or reduce the number of motor routines that must be committed to memory. (Clements, 2003, p. 328)

Rather, languages appear to exhibit economy effects over features that can best be characterized in feature-geometric terms as articulator-based nodes, such as labial or laryngeal, “which [encode] the ability of [associated] features to function as a unit in phonological processes” (2003, p. 329). Clements observes, however, that gesture economy may make more accurate predictions about phonetic repertoires, in other words, that “fully redundant features tend to be implemented in terms of uniform articulatory configurations within given classes of sounds” (2003, p. 325), such as coronal stops being realized as consistently dental or alveolar from language to language (see Dart, 1998). This is explored further in the next section, in which the distinct but related concept of uniformity is discussed.

2.1.3 Uniformity principles and “structured variation”

Research at the phonology-phonetics interface has provided us with uniformity principles as another recurring means of explanation for the relatively small number of phonetic features utilized in the typical phonological inventory. The differences between uniformity principles and economy principles are expressed by their names alone: the drive in feature or gesture economy is a minimization of storage in memory of features or gestural commands, whereas uniformity’s drive is to minimize differences among phonetic output forms if they share total or partial feature specifications. Whatever the differences between these two types of principles, however, they both similarly predict that series will tend to form and “isolated” sounds will tend to be avoided in phonological inventories. In this section, several uniformity principles are laid out in roughly the chronological order in which they are proposed in the literature.

Keating (2003), whose analysis of data from Flege (1982) is described above at length in Section 2.1.1, presents “uniformity” as a tendency for a phoneme’s various positional realizations to be produced identically in as many ways as possible; she is agnostic on whether this is strictly articulatory or acoustic in nature, or a combination of both. Keating (2003) explicitly equates this to the violable constraints encountered in Optimality Theory (Kager, 1999; Prince and Smolensky, 2004), and she refers to speakers as “ranking” uniformity above or below competing demands on speech production such as optimization with respect to context (aerodynamic, articulatory, etc.), adaptation to vocal tract morphology, and perceptual distinctness.

Keating (2003)’s notion of uniformity operates over entire phonemes: “[p]reference for uniformity amounts to a tendency to avoid allophony, that is, to let a phoneme have a consistent surface realization” (p. 377). Subsequent lines of research that independently invokes a uniformity principle generally formulate their principle as operating over smaller, subsegmental units, typically phonological features. Below, I investigate two of these accounts: Fruehwald (2013), a dissertation on the shifting vowel space of Philadelphia English, and a burgeoning line of research on structured acoustic variation in the production of obstruents

(Chodroff, 2017; Chodroff and Wilson, 2017). The latter line of work most explicitly proposes uniformity as a constraint on phonetic realization of phonological features rather than a constraint on surface similarity of entire segmental bundles of features.

Fruehwald (2013) is primarily a study of sound change in the vowel space of the population of Philadelphia English speakers. More specifically, Fruehwald discusses fronting of the long back vowels /oʊ/ (i.e., the GOAT lexical set) and /uː/ (i.e., GOOSE) and front-raising of the first part of the diphthong /ɑʊ/ (i.e. MOUTH). An important aspect of the discussion of these shifts is disentangling the effects of allophony, which creates phonological variants resistant to the broader patterns of shift. For instance, /oʊ/ and /uː/ exhibit backer variants before coda /l/ or /ɹ/ which appear to have been resistant to the fronting shift affecting other allophones over time.

There is a secondary discussion in Fruehwald (2013) of "parallel shifts" in the vowel space. Excluding major allophonic variants of a vowel, which Fruehwald argues have different phonological content from unaffected variants, he finds that the shifting formant values of particular vowels are highly correlated within their series: if one vowel is shifted in a given direction (e.g. back vowel centralization/fronting), then others in that natural class tend to shift as well. Fruehwald (2013) formalizes this observation as phonetic change targeting a given phonological natural class in a uniform way, or rather as phonetic change targeting a given feature defining that natural class in a uniform way across the segments in which the feature occurs. In terms of the population-level data he models, /oʊ/ and /uː/ front (and then un-front) in tandem in apparent time for Philadelphians.

Most recently, Chodroff and colleagues have explored the extent to which speakers are uniform in their implementation of well-understood acoustic features, under the term "structured variation". Chodroff and Wilson (2017) find that long-lag stop VOT in American English is correlated across stop place for any given speaker. In spite of considerable inter-speaker variability in the typical duration of aspiration for aspirated stops, speakers in both a single-word laboratory production task and in a multitalker corpus of spontaneous speech in a variety of rates and styles exhibit numerous significant correlations relating the duration of aspiration in their initial stops. Chodroff (2017) also reports on these experiments, but expands on this line of research considerably with a cross-linguistic survey of reported VOT values of stops in 58 languages from 24 language families. The survey confirms that the within-speaker correlations found in American English extend to a large number of other languages.

Importantly, Chodroff (2017) also moves beyond laryngeal activity as a test case for uniformity, reporting on within-speaker uniformity of the place of articulation of strident fricatives. This uniformity is evaluated using mid-frequency spectral peak (Freq_M), a spectral correlate of strident fricatives' anteriority that is relatively insensitive to prosodic and coarticulatory factors known to alter the spectral properties of fricatives (Koenig et al., 2013; Shadle et al., 2016). Chodroff (2017) finds that a large number of speakers of American English and Czech, two languages with a comparable inventory of strident fricatives (/s z ʃ ʒ/), exhibit strong correlations of Freq_M across voicing and within place, that is, between a speaker's /s/–/z/ and /ʃ/–/ʒ/ pairs. This tendency to have similar Freq_M across the voicing distinction and within the same [anterior] class is strong enough to give the impression that they are “nearly identical” within a speaker (p. 153).

Much of the theory arising from these observations is articulated in Chodroff (2017), where three uniformity constraints are proposed. The most general of these, pattern uniformity, posits a “similar pattern of phonetic targets” held in common among talkers (p. 18). As the most general formulation of uniformity, pattern uniformity allows “any template of targets provided all speakers converge on that pattern” (p. 18); in a sense, other than stipulating that speakers’ phonetic targets follow a general pattern, it allows for a large amount of unstructured variation in how phonological features are translated into phonetic targets.

The two other formulations of uniformity in Chodroff (2017) do, on the other hand, exert an influence on this mapping between phonology and phonetics. The more important for the present discussion is TARGET UNIFORMITY, which stipulates that a given distinctive feature should have a uniform phonetic value along some acoustic dimension in the various segments in which it appears. Chodroff (2017) connects this formulation to the historically common starting point in phonetics research that a consistent, one-to-one relationship exists between a phonetic target and a phonological feature (p. 20), for instance, that long-lag stops have a stereotyped timing of laryngeal articulations, and therefore a similar VOT (Keating, 2003; Cho and Ladefoged, 1999). She also explicitly equates her formulation of target uniformity with the parallelisms among vowels in a natural class observed in Fruehwald (2013)’s parallel shifts (Chodroff, 2017, p. 155).

There are two issues left unaddressed in this most recent line of work, such as Chodroff (2017) and Fruehwald (2013), that will resurface throughout this dissertation. First, it is not clear to what extent inter-speaker variability exists in the strength of the correlations among the acoustic variables measured, since data from a number of subjects is pooled in all studies involved. Inter-speaker variation in the level of uniformity, such as that discussed in (Keating, 2003), could occur in these data sets as well; variation across a population of speakers in the extent to which uniformity constrains speech production potentially carries implications for the development and change of phonological structure over time.

A second issue is that, although relatively stable, well-studied acoustic indices of articulation are chosen for analysis, articulation itself is not directly evaluated. It is therefore not immediately clear, if some types of uniformity enforce similarity or identity among targets, whether uniformity is enforced for acoustic outputs, articulatory outputs, or some flexible combination of both. One could imagine, for instance, a scenario in which a speaker produces a uniform acoustic output for a phonological feature shared across several phonemes, but implements them articulatorily in somewhat different ways; likewise, a speaker could also prioritize articulatory uniformity at the expense of uniform acoustic outputs. Both types of scenarios are occasionally attested in the experimental literature (Flege, 1982; Johnson et al., 1993; Carignan et al., 2011). This could imply that the goals of speech production may also vary analogously among speakers: some speakers may attend mainly to acoustic feedback in determining the successful production of a speech event, while others may privilege somatosensory feedback from the articulators.

To sum up, uniformity principles have, in one way or another, been a topic of discussion in phonetics and phonology for some thirty to forty years. Researchers have carried out detailed case studies of speech articulation in a number of languages (including Ewe, American English, and Czech) to test the predictions of their particular formulation of uniformity. A number of phonological primitive units have been investigated in these case studies, including at least patterns of laryngeal timing in stops, place of articulation in both strident fricatives and complex stops, and vocal tract shapes for sets of vowels. Formulations of uniformity have been proposed to hold sway over different levels of phonological representation: typically either the entire segment (Keating, 2003) or some subsegmental primitive unit such as the phonological feature and its phonetic target (Fruehwald, 2013; Chodroff, 2017) or dynamic trajectories of articulator movement that could be identified with motor programs (Maddieson, 1996).

2.1.4 Shortcomings of uniformity and economy principles

All of the economy proposals discussed in Section 2.1.2 assume that the economized unit is the phonological feature, except for Maddieson (1996). One shortcoming of these frameworks is that they discuss only the regularization of phonetic surface realization of either privative place features (Clements, 2003, and, in a sense, Maddieson, 1996) or binary features, especially laryngeal features (Keating, 2003; Chodroff, 2017; Chodroff and Wilson, 2017). At issue here is how to model the clear emergent structure in vowel systems, discussed in Section 2.1.1, which cannot be expressed easily in terms of binary or privative features, such as the within-speaker consistency in realization of multiple mid heights discussed in Ménard et al. (2008).

Additionally, it may not even be necessary for the analyst to use features to obtain inventories with structure effects that resemble the predictions of feature economy. Mackie and Mielke (2011) illustrate that vowel inventories generated without reference to a feature set as a constraining factor are still economical if the resulting vowel qualities are transcribed by experts and assigned distinctive feature values after the fact (pp. 59–62). The data tested in this study were originally generated in de Boer (2000), in which plausible vowel inventories were simulated through iterative learning undertaken by artificial communicative agents. Nor does the material being organized have to be linguistic: Verhoef et al. (2014) also observe analogous self-organization of “basic whistle elements” or “building blocks” (pp. 61–63) in the iterated imitation of slide whistle melodies.

Another concern beyond the featural architecture is that other than Martinet (1968), models of economy or uniformity are not generally motivated by non-linguistic factors, effectively reducing the proposed principles to purely linguistic desiderata. Keating (2003) formulates uniformity as a linguistic constraint, explicitly comparing it to Optimality Theory and referring to its “ranking” with respect to other linguistic constraints. Chodroff (2017) does similarly: “Uniformity should also be considered in light of *other, often conflicting constraints on the grammar* such as perceptual distinctiveness and articulatory ease” (p. 229, emphasis added).

Most interestingly, Clements (2003) compares his feature economy to structure observed at other levels of linguistic representation but does not begin to suggest that these "principles of category formation and generalisation that are at work *in other areas of grammar*" (p. 329, emphasis added) have analogues outside of linguistic organization. Given that substance-based linguistics has been formulated as a goal to "derive language from non-language" (Lindblom, 1984), this is an odd point at which to cease one's analysis. Locating domain-general principles from which these linguistic structures emerge would provide desirable connections with bodies of literature on the development of language and motoric abilities beyond language.

2.2 Basic units of articulatory reuse

The framework elaborated below differs in some crucial ways from the descriptively oriented economy and uniformity proposals evaluated in the previous section. I opt to account for the patterns described above not through a descriptive rule or constraint, but rather derive them from a tendency for a given articulatory primitive to be re-used, such that the analyst may, depending on their level of analysis, see economy or uniformity. Reflecting the multiple segmental and subsegmental levels at which uniformity may arise, uniformity is formulated in terms of continuously valued "articulatory controls", along the lines of the Perception-for-Action Control Theory or PACT (Schwartz et al., 2007; Schwartz et al., 2012). A small set of activation levels for these controls tends to be stereotyped and re-used by speakers in the process of learning.

At the most basic level lies the question of what is being re-used, in more precise terms. In the section that follows, I develop the reuse framework by considering what is to be modeled as being reused, as well as considering the circumstances under which reuse is most likely to occur. In the process, I outline a range of speech production models and the relevance of their components to the framework under development. This section develops the notion of articulatory primitive: the speech-related fields of study make available a number of primitive units of speech production that I make an effort to reconcile here.

2.2.1 Gestures and targets

The notions of gesture and target are already familiar to linguists as articulatory primitives, and I take them as my entry point for discussion. The Articulatory Phonology framework (Browman and Goldstein, 1989; Browman and Goldstein, 1992) represents speech production in terms of the discrete, purely articulatory units known as gestures. In Articulatory Phonology, gestures are defined as abstract, "cohesive unit[s]" that encompass all "task-directed movements of articulators" (Browman and Goldstein, 1989, p. 206) for a given task. Unlike the target-based models discussed below, the action units of gesturalist models are inherently connected to the individual organs of speech production (tongue, lips, larynx, velum, etc.). This has the important developmental consequence that the units of contrast are "isomorphic with units of articulation" (Goldstein et al., 2006, p. 217), grounding the process of language learning in the consolidation of speech production skill (Browman and Goldstein, 1989, p. 204).

While gestures in AP are typically described as "atomic" and implicitly indivisible, gestures as action units can in fact be decomposed into tract variables, or "independent task dimensions" that are relatively context-invariant, and sets of model articulators that act to achieve these task dimensions in a context-dependent manner (Browman and Goldstein, 1989, p. 207). Tract variables include parameters such as constriction location, constriction degree, and articulator stiffness, the latter of which determines some dynamic-kinematic properties of the gesture (Browman and Goldstein, 1989; Browman and Goldstein, 1992; Saltzman and Munhall, 1989). Tract variables are typically discretely valued, e.g. [closed], [critical], [narrow] for various constriction degrees and [labial], [coronal], etc. for various places; however, reasonable, language-specific discrete values are not always available for the constriction degree (Browman and Goldstein, 1989, p. 226) or place of articulation (Browman and Goldstein, 1989, p. 228) parameters. These have, on occasion, been modeled as continuous as a solution (Goldstein et al., 2006, p. 222). In the reuse framework developed here, mastery of a "control" corresponds to mastery of production at certain values of a coordinated set of model articulators, e.g. front raising of the tongue dorsum as in front mid and high vowels. The reuse discussed could be viewed as reuse of tract variable settings (especially continuously valued ones) across different articulator sets.

The Articulatory Phonology framework is implemented as a task-dynamic model by Saltzman, Munhall and collaborators (Saltzman and Munhall, 1989; Byrd and Saltzman, 2003), which adds substantial dynamic complexity to the system of static articulatory primitives described above. In this framework, speech production is modeled as the sequencing and interleaving of gestures with activations determined by a second-order dynamical equation that produces a critically damped oscillation about a point-attractor target. That is, the articulator set producing the tract variable represented by the point attractor approaches and does not pass the target, then returns to rest position.

I will primarily represent re-use in terms of targets, rather than gestures. Targets are logically antecedent to gestures, and are in fact contained within gestures in the form of the point-attractors used to model articulator movements toward and about targets (Saltzman and Munhall, 1989; see also Lindblom et al., 2011). Much of this model architecture is intended to model syntagmic phenomena of speech production such as timing, rate effects, and gestural co-production (Browman and Goldstein, 1992; Byrd and Saltzman, 2003). While this has led to a number of important formulations of these phenomena in terms of intergestural timing, most of this is not relevant for the present work: my interest is primarily in the content of the task variables themselves, and whether it is possible to describe a kind of paradigmatic influence among the various types of task.

Theories of speech production, and interconnected theories of speech perception, in general utilize the notion of targets at some level, whether articulatory (Fowler, 1986; Browman and Goldstein, 1989), acoustic (Lindblom, 1990; Hickok, 2014), or most commonly some combination of both (Tourville and Guenther, 2011; Houde and Nagarajan, 2011; Schwartz et al., 2012). The DIVA model (Directions Into Velocities of Articulators) is noteworthy for explicitly modeling targets as regions of acoustic and somatosensory space rather than points (Guenther, 1995; Tourville and Guenther, 2011); other models instead allow for over- or undershoot of point targets (Lindblom, 1990). These diverse characterizations of the nature of targets is in direct contrast to the fixed nature of the gesture, which is strictly articulatory in nature and modeled as a dynamical structure oscillating about a point attractor. All the same, the other models of speech production listed here and described in more detail below frequently do not measure up to Articulatory Phonology and associated task-dynamic models in terms of detailed description of the plant's task variables (here, the articulatorily grounded tract variables, assigned to specific articulator sets, that constitute gestures).

2.2.2 Neuromuscular modules and the redundancy problem

Beyond phonetics and phonology, in the study of motor control, the focus shifts to the task space in a more general sense and the problems it introduces. The number of degrees of freedom typically exhibited by a physical plant is prohibitively large for direct control of each degree of freedom by the central nervous system; moreover, many combinations of plant settings along available degrees of freedom will result in the same task outcome (Bernstein, 1967). The speech articulation apparatus is particularly notable for its large number of degrees of freedom, both in the sense of the structure and potential activity of the musculature (Sanders and Mu, 2013) and in terms of the large range of outputs that can be produced through this action, which can be situated along a very large number of potentially perceptible somatosensory and acoustic dimensions (Hickok et al., 2011; Gick and Derrick, 2009; Gick and Stavness, 2013).

For computation of appropriate motor controls to be handled, realistically, models must reduce the degrees of freedom of control over the plant to a tractable level; this is sometimes called the REDUNDANCY PROBLEM or DEGREES-OF-FREEDOM PROBLEM. Analysts have primarily argued for two types of simplification. The first is organizing control of the plant into functionally defined groups of muscles whose correlated activations are directed towards particular tasks. This type of abstract coordinative unit was first referred to as the neuromuscular synergy (Bernstein, 1967) and it is still frequently referred to as such in the motor control and neurolinguistic literatures (see, for instance, d'Avella et al., 2003; d'Avella and Bizzi, 2005; Rugy et al., 2013; Sohn and Ting, 2016). The term module (Gick and Stavness, 2013; Berger et al., 2013) or coordinative structure (Easton, 1972; Turvey, 1977) is occasionally encountered instead, the latter often in reference to problems in speech production (Fowler and Turvey, 1980; Kelso et al., 1984). Modular or synergistic units can often be identified with gestures (Gick and Stavness, 2013; Ramanarayanan et al., 2013); this isomorphism among gestures and a range of other models outside of linguistics proper is a factor in my using gesture-oriented representations to elaborate the reuse framework here.

The other means of solving the degrees-of-freedom problem is to incorporate an optimization algorithm into the speech production model: that is, the potentially infinite solutions to the degrees-of-freedom problem are solved by choosing the one solution that minimizes some costs given a particular set of starting conditions and internal and external feedback. These optimal control strategies constitute the consensus view of the algorithmic basis of motor activity (Todorov and Jordan, 2002; Todorov et al., 2005; Loeb, 2012; Friston, 2011) in most fields adjacent to speech science. The use of synergies and optimal control strategies are not mutually exclusive, but both solve the same "degrees-of-freedom" problem of control over the plant. Because modular or synergistic organization of the plant's degrees of freedom does reduce the controllable degrees of freedom to a manageable level, there are some implications for the choice of overall control style, the two primary versions of which are explored further in Section 2.3.

2.2.3 Generality of speech motor learning and reuse

Another parameter of articulatory primitives with implications for the reuse framework developed here is whether reused articulatory primitives have some connection to global control parameters. Systems of related goals that use the same plant have been described as using a global sensorimotor map for motor control purposes. The primary means of investigating this type of globally recruited motor control element is to examine whether adaptation to perturbations on one task easily generalizes to another related task, or whether no such transfer of behavior occurs. In the latter case, the learning is said to be specific to particular task contexts. Put another way, the nature of a given target and other targets that it may be uniform with is unclear: do the related production goals of these several sounds derive in part from a single shared control structure, thus literally reusing control elements? Or are multiple, task-specific control mechanisms each associated with a specific speech sound and merely constrained in their similarity to one another (see Chodroff, 2017)? The choice of one model or another has implications for the architecture of the reuse framework developed here.

One line of research on learning and generalization of motor adaptations from one set of sounds to another, most frequently vowels or sequences of vowel targets, finds a "broad but decaying pattern of generalization" of learned adaptations in the production of one segment to other segments (McFarland et al., 1996; Houde and Jordan, 1998; Villacorta et al., 2007; Cai et al., 2010; Thibeault et al., 2011). The factors contributing to the limited specificity that is observed (the "decay" in generalization) vary from study to study: evidence is found for specificity to vowel formant targets (Houde and Jordan, 1998; Villacorta et al., 2007) and the sequencing and dynamics of multiple vowel formant targets (Cai et al., 2010). Regardless, there is still some effect on a given segment exerted by adaptation to perturbations made to some other segment, suggesting that auditory-motor mappings used for controlling the plant encode entities that are shared among vowels, possibly due to shared reference to a global controller or sensorimotor coordinate map (Cai et al., 2010, p. 2046).

Other studies find substantially less generalization and on that basis argue that speech motor learning is fundamentally specific to single targets or sequences of targets (Tremblay et al., 2008; Pile et al., 2007). A theme that emerges from this line of research and its interpretation is that "patterns of generalization [are] highly dependent on the nature of the task" (Malfait et al., 2005). The minimal training regimes used in these experiments may explain the large difference in results compared to the group of papers discussed above: most critically, the number of distinct lexemes that study participants were trained on (and which were subjected to the perturbation) differs substantially among studies. Experiments in which generalization is observed train on multiple lexemes containing the phones at issue (Cai et al., 2010; Villacorta et al., 2007), while experiments where little to no generalization is observed most often train on a single lexical item repeated across all training trials (Tremblay et al., 2008; Pile et al., 2007). Tremblay and colleagues note that "[g]eneralization in speech learning would presumably be evident after training with multiple utterances that could be combined to form the elements of the transfer set" (2008, p. 2432).

2.3 Internal models and model-free learning mechanisms

In this thesis, I treat reuse or uniformity as resulting from general properties of motor learning, rather then from a purely linguistic desideratum. The mechanisms that lead to regularities in phonetic and phonological structure of the sort addressed by economy or uniformity principles are, in the framework I develop here, inseparable from the process of learning and consolidating speech motor programs. In this section, I discuss several accounts of internal models, or the implicit knowledge of the somatosensory and auditory outputs that result from applying given motor controls to the plant, and their role in motor learning. In terms more familiar to linguists, an internal model is essentially a detailed motor-articulatory-acoustic mapping. Assumptions about the nature of internal models and their involvement in motor learning vary among models of speech motor control, although the study of model-based learning predominates in the motor control literature (Haith and Krakauer, 2013). In this type of learning, which is broadly implicated in implicit adaptation to changes in task-relevant conditions, an internal model is responsible for recalibrating the controller to apply controls appropriate for task achievement (Section 2.3.1).

Developing an internal model that can be used to alter a task's controller in a way that facilitates successful task completion requires considerable time and effort, and presumes a long period of gathering information on the dynamics of the plant and how it responds to motor inputs. Model-based learning is therefore not feasible as a strategy during L1 acquisition, at which time the learner has not gathered enough information on motor-articulatory-acoustic mappings. In Sections 2.3.2–2.3.3, I introduce a distinct type of learning strategy, variously termed explicit, model-free or cognitive strategies (Haith and Krakauer, 2013; McDougle et al., 2016), that may predominate during L1 acquisition and other low-information learning situations, particularly in the form of trial-and-error-based selection of articulatory strategies. In Section 2.4, I review the evidence that hallmarks of these strategies, including articulatory reuse, can be seen during L1 acquisition.

2.3.1 Model-based learning: state estimation and state feedback control

Model-based learning, which is implicated in adaptation learning of the sort described above, predominates the motor control literature, including the development of its most popular model architectures. A central feature of most models of motor control is the use of internal feedback, rather than external (sensory) feedback alone, to compare expected and desired somatosensory and auditory consequences of an action in progress. The expected consequences are predicted with a process of state estimation, in which an internal model or models map between a plant's current state and and its subsequent state, given the currently active controls and the state of the plant (Miall and Wolpert, 1996; Houde and Nagarajan, 2011). The computations carried out by these internal models are often modeled as a Kalman gain function, also known as a Kalman filter (Wolpert et al., 1995, p. 213; Todorov and Jordan, 2002, p. 1234), and are often localized in the cerebellum (Wolpert et al., 1998; Taylor et al., 2010; Manto et al., 2012) and cerebral cortex (Houde and Nagarajan, 2011; Hickok, 2014). Internal models solve important computational problems for motor control: using an internal forward model rather than sensory feedback alone allows the controller to compare sensory information and expected dynamical conditions of the plant faster than external sensory information on the plant's status would reach the central nervous system (Miall and Wolpert, 1996; Wolpert and Flanagan, 2001; Houde and Nagarajan, 2011).

Models using some type of internal state estimation as feedback include optimal feedback control (Todorov and Jordan, 2002; Todorov et al., 2005). Optimal feedback control is achieved by minimizing a cost function at each time step during task-directed action, thereby finding the optimal controls to apply to the motor effectors in use to achieve the desired outcome. This section primarily focuses, however, on State Feedback Control (SFC), another feedback control model that has been elaborated both for motor control generally (Guigon et al., 2008; Shadmehr and Krakauer, 2008) and for speech motor control specifically (Hickok et al., 2011; Houde and Nagarajan, 2011; Hickok, 2014).

As described in some detail in Houde and Nagarajan (2011), SFC models as applied to speech motor control posit a particular type of state estimator as the source of internal feedback. The state estimator contains linked forward models that first predict the upcoming dynamical state of the plant, given the current state of the plant and efference copy of the controls currently being applied; the estimated dynamical state is then used to predict the sensory consequences (auditory and otherwise) of the estimated plant state. Actual sensory outputs (external feedback) are then compared with the predicted outputs (internal feedback), generating SENSORY PREDICTION ERROR signals. Based on these mismatches between predicted and actual feedback, corrective controls are then generated using a Kalman gain function.

In SFC models, computation of internal feedback, which assumes the existence of an internal model of the sort described above, resides squarely in the central nervous system (Houde and Nagarajan, 2011), rather than in the lower nervous system, as earlier models such as equilibrium point control have located it (Perrier et al., 1996; Sanguineti et al., 1997). An extension of SFC, hierarchical state feedback control (HSFC), has two levels of state feedback control (with two pairs of controllers and predictors) implicated in the task of speech: a high-level controller for syllable-level program "chunks" and a lower-level controller for detailed attainment of the targets passed to it (Hickok, 2014). As alluded to above, however, the inventory of targets at either level is not a focus of elaboration in descriptions of SFC and HSFC, which instead emphasize the model's architecture and the neuroanatomical localization of its various components.

The Directions into Velocities of Articulators (DIVA) model is another commonly encountered feedback control model with a long history in speech science. The DIVA model's architecture is comparable to the SFC models described above, with some differences. External feedback correction of controls generated by the forward model occurs specifically when sensory error indicates that output has crossed a perceptual boundary into another phoneme (Guenther, 1994; Guenther, 1995; Guenther et al., 2006; Tourville and Guenther, 2011). The internal forward model itself is not modularized as in SFC, but an inversion of sensory feedback error to articulatory corrections is accomplished via a learned mapping to articulator movements from orosensory and auditory feedback (Guenther, 1995, p. 50), which creates a similarly difficult computational task for the CNS. One distinguishing advantage of the DIVA model is that the inventory of tract degrees of freedom manipulated by the model's controller(s) are substantially more elaborated than in the models described above: the orosensory dimensions and articulatory degrees of freedom that comprise DIVA's task dimensions are in fact similarly detailed to Saltzman and Munhall (1989)'s tract variables (Guenther, 1995, p. 60). However, the DIVA model does exhibit a number of differences from the dynamical system framework. Production is a smoothly connected string of targets, which are modeled as convex regions in acoustic space (Guenther, 1994), rather than a dynamical system based on coupled oscillators.

DIVA's elaboration of tract variables does come with some drawbacks associated with the model. The DIVA model's neuroanatomical implementation of its neural network, which in early versions has one node that activates per distinctive sound during perception and production (Guenther, 1995), clashes with findings showing somatotopic organization of speech motor cortex during speech production (Bouchard et al., 2013; D'Ausilio et al., 2014). The other major drawback is that DIVA's trajectory corrections (generated by the central nervous system) do not take dynamical state into account as is the case for SFC and HSFC, but rather leave the control of articulator dynamics to the lower motor system (Houde and Nagarajan, 2011), in spite of experimental evidence suggesting that the CNS does directly handle corrections to plant dynamics (Tremblay et al., 2008; Nasir and Ostry, 2009; Cai et al., 2010).

To sum up, model architectures for speech motor control most frequently involve a well-developed internal forward model that provides internal feedback to the controller via a state estimation process. Some variation is seen in the computational basis of the recalibrations of the controller (i.e., optimal feedback control versus state feedback control). The parameters of vocal tract motor control are not typically explicitly spelled out in great detail, but are elaborated in the DIVA model and in the task-dynamic models of speech production that accompany Articulatory Phonology. The details of a control architecture ideal for this thesis are easy to sketch out: an internal model updates controls applied to gesture-like coordinative structures. The isomorphisms between DIVA's task variables and the gestural primitives in Articulatory Phonology, then, are crucial for the discussion at hand in that they link the desired motor control parameters with the desired controller.

However, the above models of speech motor control for the most part account for model-based learning of adjustments in already-learned task spaces, such as L1 speech by typical adult speakers. Speech production is a complex collection of interrelated tasks that takes a notably long time to master, in terms of producing speech in the same fashion as the surrounding community. The development of an internal model is also non-trivial; in more familiar terms, the motor-articulation and articulation-acoustics mappings must be learned, and the learner must expend a great deal of effort to do so. In the following sections, I consider a different motor learning strategy based on trial-and-error exploration and reuse during action selection, which may predominate in cases where the internal model has not fully developed and state estimation is not yet fully reliable.

2.3.2 Explicit selection and trial-and-error learning

Although less heavily researched than learning driven by internal models, so-called cognitive or strategy learning processes have also seen some research. Strategy is thought to be a mechanism of learning distinct from adaptation (discussed in Section 2.3.1) in timescale, cognitive involvement, and sources of feedback (Haith and Krakauer, 2013; McDougle et al., 2016). Crucially for the argument advanced in this dissertation, strategy-based learning has also been argued to be model-free in some cases; that is, not involving an internal model, with the learner instead directly updating the controller through explicit selection of controls to be applied. While there is a notable dearth of studies on strategy learning in speech production, some generalizations can be gleaned from the broader motor control literature and are reviewed here.

A wealth of experimental evidence suggests a qualitative division of learning into at least two types of components: explicit components driven by strategy and implicit components driven by internal models (Shadmehr et al., 2010; McDougle et al., 2016). Strategy learning is based on explicit decision-making, may result in the fast discovery of local optima of performance (Bond and Taylor, 2015; McDougle et al., 2015), and the timecourse of performance improvements is non-monotonic (Taylor et al., 2014; Bond and Taylor, 2015). This component is clearly separable from adaptation driven by an internal model, primarily by the latter's monotonic improvement over time and relative lack of flexibility (Taylor and Ivry, 2011; Taylor et al., 2014; McDougle et al., 2015; Huberdeau et al., 2015; Sülzenbrück and Heuer, 2009). The two learning processes are in fact often at odds with one another in terms of task achievement, since implicit learning and recalibration of control policy will carry on even if it begins to interfere with a successful strategy discovered through trial and error (Mazzoni and Krakauer, 2006; Taylor and Ivry, 2011; Benson et al., 2011; McDougle et al., 2016).

The two types of learning also rely on feedback provided by different sorts of error. PERFORMANCE ERROR, or the difference between a goal and an observed outcome (Taylor and Ivry, 2011), is used in strategy learning, and is clearly distinct from sensory prediction error, the difference between estimated and actual sensory feedback, which has been discussed above as an important aspect of feedback control models (Shadmehr et al., 2010; Houde and Nagarajan, 2011). The point of comparison for the observed outcome is different: rather than a state estimator's implicitly understood expected outcome, performance error compares the observed outcome to a goal, which may explicitly comprehended by the learner. Accordingly, explicit learning (using strategy) may involve active exploration of the task manifold or repeated hypothesis-testing (Haith and Krakauer, 2013; Wilson et al., 2014; Loeb, 2012), in line with the non-monotonic progress described above.

Cognitive, explicit learning is most likely the type of learning that predominates during speech motor learning, particularly during language acquisition. It stands to reason that if sensory outputs are unreliable or cannot (yet) be reliably interpreted in terms of motor inputs, model-based learning mechanisms cannot predominate as they do in speech motor control in healthy adults. Cognitive strategies are known to predominate in populations whose ability to maintain an internal model is reduced due to cerebellar degeneration (Taylor et al., 2010); it seems reasonable to assume this learning style for neurologically normal but still-developing language learners.

Crucially, strategy learning has been argued to be MODEL-FREE in many instances: not making use of an internal model that assists in implicit corrections based on feedback, and instead using an explicit process of trial and error to identify successful strategies (Haith and Krakauer, 2013; Wilson et al., 2014) and directly modify the appropriate control policies. This explicit exploration process is one solution to the so-called explore-exploit dilemma in learned behaviors: exploration of less well-understood portions of the task manifold in pursuit of a greater performance optimum versus continuing to exploit a local performance optimum that is sufficient, but not globally optimum, for task achievement (Wilson et al., 2014). Exporation is presumed to proceed in terms of selection of a set of activations of primitives (i.e., synergies) from the task-equivalent manifold (Diedrichsen and Kornysheva, 2015), followed by execution of the selected strategy and evaluation of the resulting performance error.

If model-free, explicit learning predominates during language acquisition, activations of task variables (muscular synergies, gestures, etc.) known to result in success at one task may be deliberately extended to new tasks with comparable goals (i.e. VOT timing across stops as they differentiate), thus "yoking" together articulatory strategies through common task variable activation patterns. These shared or reused activations of task variables may constitute the neural substrate of articulatory uniformity as it is discussed above. In the next section I consider additional theoretical work that clarifies this association between articulatory reuse and the explicit selection of articulatory controls during trial-and-error learning.

2.3.3 Good-enough control: an alternative model centered on reuse

Still another account of motor control suggests a more radical reconsideration of how motor learning is typically achieved. Here, I highlight an especially interesting proposal by Loeb (2012): that the fully detailed internal models described above (particularly in Section 2.3.1), which relate sensory error signals to estimated plant states, are implausible. In particular, solving the degrees-of-freedom problem by minimizing a global cost function at each time step during action is said to involve too much computation (Loeb, 2012, p. 758). The alternative suggestion is that control in biological organisms is merely GOOD ENOUGH, rather than optimal. This proposal not only has intriguing parallels with the accounts of cognitive or explicit learning reviewed in the previous section—in that it does not implicate an internal model—but it also contains some elements that make it compatible with existing linguistic exemplar theories.

Good-enough control is intended as a computationally and phylogenetically feasible model of motor control in biological organisms. Under good-enough control, application of the appropriate controls to the plant is not a product of cost function minimization or state estimation as described in Section 2.3.1, but rather results from a storage-and-lookup process associated with motor areas in cerebral cortex. Loeb (2012) argues that storage and lookup of event traces, rather than performing complex computations, is more likely the function that human cerebral cortex has evolved to carry out. As such, controls are applied to a given task situation as the result of a "search" for stored instances of past motor activity similar to the current situation, sometimes followed by "minor interpolations and adjustments" to account for new conditions or a slightly different task (Loeb, 1983; Loeb, 2012, p. 760). For the purposes of this dissertation, a characteristic behavior associated with Loeb's good-enough control, the "hacking" of one set of controls to generate a novel one (Loeb, 2012, p. 759), may be identified with articulatory reuse as discussed further below.

Good-enough control is argued to account better than optimal control for certain behaviors surrounding the selection, execution, and refinement (learning) of motor controls. At the level of human performance on motor tasks, the storage-and-lookup approach of good-enough control and the cost function minimization approach of optimal control may be indistinguishable from an internal model while being more computationally realistic for the neural substrate involved (Loeb, 2012, p. 759). But good-enough control explains broader behavior better, particularly in systems for which the task-relevant controls are not already mastered. The learning curve on tasks for which novel controls must be mastered resembles a random walk for individual subjects, which often averages to a smooth, monotonic improvement in studies with large numbers of subjects (Gallistel et al., 2004). This pattern is better explained by good-enough control's patterns of exploration of the plant's control space and opportunistic retention of successful motor strategies (Loeb, 2012, p. 762). This opportunistic retention and commitment to memory of controls that happen to work also may relate to observed persistence of motor habits, even when these habits are suboptimal (Diedrichsen et al., 2010; Rugy et al., 2013; Sohn and Ting, 2016).

Good-enough control may also be appealing to linguists for other reasons. Good-enough control replaces the learned internal forward model typical of other speech production models with a stored repository of past task completions that can be compared with the current goal. There are some similarities between this storage-and-lookup component of the model and the mechanisms underlying exemplar models of speech perception and production (Johnson, 2006; Pierrehumbert, 2008), which similarly involve storage of episodic traces and reference to these traces during perception and production. Such models in phonetics could easily be expanded to include storage of articulatory strategies and neuromuscular activations, along with associated evaluations of success.

2.4 Articulatory reuse as a consequence of trial-and-error learning

Motor learning cannot be modeled without some implicit learning processes and some role for internal models. There is a great deal of experimental evidence for learned internal forward models in motor control generally (Wolpert et al., 1995; Miall and Wolpert, 1996; Shadmehr and Krakauer, 2008; Tian and Poeppel, 2010), and most theoretical discussions of model-free, explicit selection of strategy are quick to point out that both mechanisms are involved in most learning processes (Haith and Krakauer, 2013; McDougle et al., 2016). In particular, in speech motor control in adults, there exists evidence that internal feedback is readily, rapidly incorporated into control of speech (Shadmehr et al., 2010; Houde and Nagarajan, 2011; Niziolek et al., 2013). Model-driven, implicit learning as a part of motor control is simply too useful an idea to abandon entirely.

Nonetheless, model-free, explicit, exploratory learning processes may dominate implicit, model-based learning processes during language acquisition. The existence of a highly reliable internal model and its associated components to learning are well-supported for adult populations, but not necessarily developing speakers, who may lack detailed knowledge of acoustic-articulatory mappings for their L1 or for portions of their L2 (or L3, etc). The acquisition of speech motor control in fact resembles nothing more than a prolonged series of explorations of the (undeveloped, growing) task space for speech articulation. Under these conditions, articulatory reuse has a simple motivation: conservation of energy and relatively quick achievement of some external reward by "hacking" is much easier than "developing a new sensorimotor program from scratch" (Loeb, 2012, p. 761), particularly for child learners.

In this section, I review the evidence for dominance of explicit, model-free learning strategies during early language acquisition. I place particular emphasis on evidence for reuse of stereotyped articulatory strategies in child language, identifiable with an extreme version of the process of "hacking" as discussed in Loeb (2012). Provided that reuse during learning occasionally leaves traces in adult language in the form of covariation of acoustic or articulatory attributes of segments, articulatory uniformity would then arise not from a specifically linguistic demand but from a limiting factor on all motor learning.

2.4.1 Articulatory reuse in child language

Babbling is generally acknowledged to begin the development of the child's internal model (Locke, 1983; Guenther, 1994; Vihman et al., 2009; McAllister Byun et al., 2016). An optimal control system with fully-formed internal models would be expected to show monotonic improvement in achievement of task goals. From the "first words" stage and onward, as more complex sequences of speech motor tasks are demanded, however, the developmental trajectory of children's speech is far from gradual linear progress. Following an often-accurate initial attempt at adult-like production, children most often engage in wide-ranging exploration of the articulatory possibility space, with some regression to stable but inaccurate strategies (a "U-shaped" attaniment curve) often occurring before ending in accurate, stable outputs, as schematized in Figure 2.3 (Menn, 1983; Becker and Tessier, 2011; McAllister Byun et al., 2016). As the child explores the articulatory possibility space, a good-enough control strategy may dominate (or trial-and-error selection may predominate), during which the capacity for speech production is restricted to a relatively small repertoire of motor programs learned by chance success. The dominance of lexical selection and templates in early development (see below) suggests that children expand this early repertoire via trial-and-error extension of old controls to new contexts.

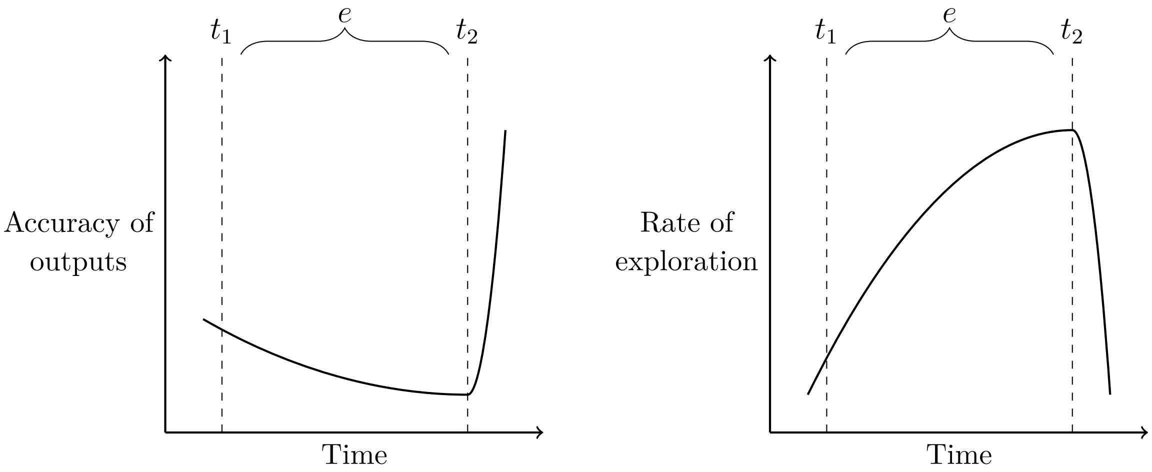


Figure 2.3: Schema of the "U-shaped" attainment curve in development of speech production capacity as it relates to exploration of the articulatory task space. The time point t_1 refers to initial attempts at producing an imitation of the adult model; e refers to a period of intensifying, random exploration of the speech articulation task space; and the time point t_2 refers to the eventual selection of an applicable articulatory strategy that produces precise, accurate outputs.

Evidence for a developmental contribution to phonetic and phonological structure regularization can be found in a range of research on normal and disordered phonological development in children. Child learners extensively re-use their mastered articulatory strategies, which can be modeled as heavy exploitation of a small set of "stored production subroutines" to imitate adult models (Menn, 1978, p. 164; Menn, 1983). This reference back to a small set of learned subroutines is more computationally efficient than continuously inventing new subroutines to better imitate the adult model (Menn, 1978, p. 165; Menn, 1983, pp. 30–32), with parallels to Loeb (2012)'s observations on development of new motor controls. Experiments on imitation also show that children more accurately reproduce sequences in nonce words when those sequences are already frequent in their lexicons; this effect is stronger for children with atypical phonological development (Beckman and Edwards, 2000). Generalization has also been noted of newly practiced strategies for producing [+voi], [+continuant], and [+strident][3] to untrained phonemes containing these features (McReynolds and Bennett, 1972).

Early speech production experience, most likely acquired using trial-and-error learning, aggressively generalizes to new words once word production becomes a major goal of the child learner, with learner-internal consistency often winning out over resemblance to the adult form. The child's idiosyncratic language-learning experience systematically affects uptake of lexical material from the environment: words that contain mastered articulatory patterns are learned at a greater rate than words that do not (Ferguson and Farwell, 1975; Menn, 1983; Vihman and Velleman, 2000; Vihman, 2014). Stereotyped phonotactic structures known as templates are also well-known features of early language development (Menn, 1983; Vihman and Velleman, 2000; Vihman and Croft, 2007; Vihman, 2014); these are said to result from idiosyncratic "systematization" of adult inputs according to the articulatory sequences that happen to be already mastered (Vihman and Velleman, 2000). That child language learners generalize a handful of successful motor routines in this fashion in the course of building a larger lexical (and, presumably, gestural) repertoire suggests that a "good-enough" regime dominates at this time period, where the compromise implicit in that term is in terms of similarity to adult targets in the ambient language environment.

Research in child phonology by McAllister Byun et al. (2016) also strongly suggests a primary role of "hacking" and associated trial-and-error learning strategies during the consolidation and refinement of phonological targets, as would be expected under a regime of good-enough(-like) control. In their A(rticulatory)-Map framework, child language learners are restricted in their ability to use sensory feedback. According to McAllister Byun et al. (2016), learners have an articulation-acoustics mapping, but it is underdeveloped given the small number of reliable motor programs in any child learner's repertoire (p. 150). This limited internal model (the A-Map) is stated to be the same sort of model as the learned set of forward and inverse calculations posited for optimal control (Miall and Wolpert, 1996; Wolpert, 1997; Wolpert et al., 2001) but in fact differs strikingly. The A-Map rather takes an exemplar-theoretic approach that recalls Loeb (2012)'s storage-and-lookup module (McAllister Byun et al., 2016, p. 142) instead of the computation typical of optimal control models.

---

[3] Interestingly, McReynolds and Bennett (1972) employ the SPE definition of [strident] that extends to labiodental fricatives (Chomsky and Halle, 1968), and the generalization effects observed for [+strident] are said to improve productions of sibilants and labiodentals alike, but not the non-strident interdental fricatives such as /θ/.

The result of the A-map architecture is that precision (ability to hit approximately the same target, given low motor ability) is prized over accuracy to adult targets, seemingly opening the door for reapplication of already-consolidated controls to new speech production tasks.

2.4.2 Traces in adult language

While further direct consideration of child language is beyond the scope of this dissertation, a pivot to adult language is not without its merits, as I argue below: adult language is not independent from the starting state provided by that learner's experiential idiosyncrasies during language development. In fact, the idiosyncratic but consistent-within-speaker results of "hacking" during trial-and-error selection, which predominates during language acquisition, may result in the constrained inter-category phonetic variation eventually observed in adults.

Ménard et al. (2008) arrive at such a conclusion based on experimental evidence of covarying vowel qualities for both child (ages 4 and 8) and adult speakers of European and Québecois French. Production of vowel height in the two degrees of mid vowel in French (mid-high /e, ø, o/ versus mid-low /ɛ, œ, ɔ/) is constrained to a high degree of similarity: speakers produce the vowels in each set with statistically indistinguishable F1 values. Ménard et al. (2008) provide an analysis within the Perception for Action Control Theory (PACT) framework (Schwartz et al., 2007; Schwartz et al., 2012). PACT posits a multimodal speech percept, incorporating somatosensory information on both articulation itself and its acoustic results into stored representations of speech events. Ménard et al. (2008) conclude that the observed covariation of acoustic vowel qualities in a broad range of speakers, including children, is a consequence of bootstrapping the production of all vowels of a given height off of a single vowel learned first, most likely the front unrounded vowel in each set.

Analogous findings in research on articulatory uniformity, briefly reprised here from Section 2.1.3, may also be taken as evidence for this traces of "hacking" left over from acquisition. Long-lag VOT at the full range of stop places is implemented with considerable similarity within-speaker (Keating, 2003; Chodroff and Wilson, 2017). It stands to reason that a child learner first learns how to produce a single set of under-differentiated stops (bilabial or linguopalatal) and then extends these timing controls to new contexts, resulting in the "yoked" adult production strategies seen here and in (Ménard et al., 2008). Analogues can be imagined for many of the other uniformity findings: fricative place is remarkably similar across voicing, as indexed by mid-spectral peak Chodroff (2017), and if one American English back vowel is acquired with a relatively front, diphthongized value, the others tend to be as well (Fruehwald, 2013).

Some models of speech production and perception include a factor that regularizes articulatory strategies used by individual speakers during language development (Lindblom, 1998; Studdert-Kennedy, 1998). This factor is typically connected to the relative ease of re-using an articulatory routine in place of learning another control, with clear similarities to Loeb (2012)'s "hacking". For instance, Lindblom (1998) includes a component in his model of phonological development that amounts to “phonetic nepotism”: it “improve[s] the learnability of all forms having the same movement components”, regardless of the level of articulatory difficulty for that form (p. 259; see also Lindblom et al., 2011, pp. 86–88). This approach to language learning accords with the “particulate principle of self-diversifying systems”, which dictates that combinatorial systems like language are based on simpler, recombinable “particles” (Abler, 1989). Some further theorize that these “particles” are targets for learning (Studdert-Kennedy, 1998; Goldstein and Fowler, 2003; Goldstein et al., 2006); these are often explicitly recognized as gestures, said to originate in the learner’s “recruit[ment] [of] well-practised action units” for the task of speech (Browman and Goldstein, 1989).

2.5 Conclusion

In this section, I offer some concluding thoughts. Section 2.5.1 summarizes the relationship between the various threads of research on motor learning and linguistic structure and motivates the experiment described in Chapters 3 and 4. I elaborate upon the specific need for further research on articulatory uniformity in vowel systems using articulatory imaging in Section 2.5.2.

2.5.1 Directions for research

This chapter has made an effort to connect three broad areas whose research goals are not well-integrated at present: the general motor control and learning literature, research on language development, and substance-based phonology, including research on uniformity principles and “structured variation”. Two areas of particular joint concern are discussed above. On one hand, patterns inherent to model-free learning appear to influence the selection and reuse of articulatory strategies during language acquisition. On the other hand, the reuse of articulatory strategies during language acquisition appears to constrain variation of phonetic implementation of phonological primitives for adult speakers. Both areas require further investigation to ensure that behavior is actually congruent with the model laid out here.

One consequence of the articulatory reuse discussed above is the articulatory uniformity of lingual postures in sounds with comparable acoustic goals, but where the uniformity need not exist for contrast to be maintained: uniformity for uniformity’s sake. The experiment in Chapter 4 examines a population of Sūzhōu Chinese speakers for uniform production of fricative consonants and certain vowels that appear to have developed fricative noise targets in the recent past. Finding uniformity in this situation would add to evidence consistent with the model laid out above: that reuse arising from exploration during model-free learning may leave an imprint on the adult inventory. Sūzhōu Chinese speakers, during languge acquisition, may stumble upon the strategy of producing vowel sounds with fricative noise with the same lingual posture as the most similar fricative, even when those same tongue shapes do not historically characterize the vowels in question.

Research on uniformity in vowels, specifically, is important because it extends the study of uniformity beyond the timing relations it is historically founded upon. The timing of VOT in long-lag and short-lag stops has been the focus of the bulk of studies of uniformity (Keating, 2003; Chodroff and Wilson, 2017; Chodroff, 2017). Demonstrations of uniformity are of limited usefulness if they are only observed in such a small portion of the speech articulation capacity, particularly since these observations mainly concern uniformity in timing relations among articulatory events and not uniformity of the vocal tract shape primitives used in the events themselves. There are some exceptions to this tendency, such as Chodroff (2017)'s work on uniformity of fricative anteriority contrasts across voicing contrasts. Other studies occasionally provide incidentally useful data that were not collected for the specific purpose of demonstrating uniformity (i.e., Dart, 1998) and so are less specifically useful.

Vowel production involves control of a substantially larger number of degrees of freedom than stop production. Vowels, as well as liquids, thus afford the opportunity to investigate relations among several independently controllable articulatory primitives, such as tongue dorsum raising, tongue convexity, and lip rounding; as well as several controls which cross-cut other controls (i.e., height as realized in front and back vowels). Development of experimental paradigms for studying uniformity in vowel production is critical because evidence for uniformity across a wider range of acoustic or articulatory dimensions—or its absence in particular dimensions—would help to test the predictions of a theory of uniformity.

2.5.2 The importance of articulatory measures of uniformity

The articulatory-acoustics mapping for oral vowels is thoroughly researched and well understood (Fant, 1960; Lindblom and Sundberg, 1971; Wood, 1986; Stevens and Keyser, 1989; Johnson et al., 1993; Ménard et al., 2008). One may therefore question why the means of measuring uniformity in the experiment in Chapter 4 are articulatory rather than acoustic. I use this concluding section to justify the use of articulatory data in this dissertation, rather than simply investigating known acoustic correlates (formant frequencies) of the major articulatory vowel features (height, backness, rounding, etc). This is critical because the (SC) many-to-one problem opens the door to functionally similar outputs having substantially different inputs, many of which may reveal covert uniformity or non-uniformity in theoretically interesting ways.

Task-directed speech motor activity is highly dependent on the individual, down to idiosyncratic combinations of muscular activations for relatively simple tasks like opening the jaw (Gentil, 1992). In the more complex tasks typical of speech production, it is clear that idiosyncratic differences are present at all levels of articulatory strategy, from the single-articulator strategies associated with specific categories (Johnson et al., 1993; Westbury et al., 1998; Zhou et al., 2008; Noiray et al., 2014; Mielke et al., 2016) to the so-called "trading relations" that hold among articulators (Perkell et al., 1993; Guenther et al., 1999). These idiosyncrasies could well encompass some uniformity of articulatory strategy that is otherwise masked by many-to-one mappings between articulation and acoustics (Atal et al., 1978; Maeda, 1990).

A focus on the acoustic correlates of articulation alone thus likely understates the extent to which speech articulation is uniform for a given individual. I reprise discussion of the data from Flege (1982) to demonstrate this point (see Figure 2.1). Laryngeal timing is thought to have a relatively direct relationship between articulation and acoustics, particularly in the area of voice onset time. However, as seen in Flege (1982), speakers exhibit considerable idiosyncrasy in their articulation-acoustics mappings even for this well-understood articulator. For instance, prevoicing is somewhat unpredictable from vocal fold adduction, which is thought to be the primary determinant of whether acoustic prevoicing occurs. In a sense, then, this data set presents a worst-case scenario for those who insist that unique one-to-one mappings from a given articulatory configuration to a single acoustic output must hold. More direct measurements of articulatory activity are needed to fully evaluate the extent to which uniformity applies for a given individual and set of segments.

Chapter 3

An overview of Sūzhōu Chinese and its fricative vowels

Vowel systems offer an interesting opportunity to investigate articulatory reuse. Vowels exist in a relatively continuous possibility space, in contrast to the relatively discrete constriction locations of consonants, and application of similar motor controls to the articulatory apparatus intuitively connects series of vowels sharing a linguistically relevant feature (height, anteriority, rounding, etc). These controls must be applied across different "contexts" within the series, however: a front tongue position may be combined with a variety of target jaw heights, lip apertures, or convex or concave postures of the tongue. Compelling evidence for articulatory reuse is found when these controls are applied in precisely the same way across multiple sounds (Ménard et al., 2008), but it is often difficult to make the comparison.

Some articulatory controls may be shared across a set including consonants and vowels, such that resemblance of a consonant and a vowel can be taken as especially strong evidence of articulatory reuse at some point during the speaker's lifespan. FRICATIVE VOWELS, so called because they are identifiable as phonological vowels using language-specific diagnostics but exhibit fricative noise production targets, make a compelling case study. These fricative noise targets, which are produced using supralaryngeal constrictions (i.e. not using non-modal phonation), may or may not share articulatory strategies with acoustically similar fricative consonants in the phonological inventory. Fricative vowels are attested in small pockets globally, including in dialectal Swedish (Bruce, 2010; Westerberg, 2016) and Grassfields Bantu (Faytak, 2017); they are found especially often in Chinese, especially Mandarin and Wú 吴 Chinese dialects (Qian, 1992; Zhu, 2004; Zee and Lee, 2007).

Sūzhōu 苏州 Chinese, a Wú dialect closely related to Shànghǎi 上海 Chinese, is of interest here for the specific articulation of two of its fricative vowels, which exhibit a postalveolar constriction and are transcribed here as /ı͡ʑ/ and /y͡ʑ/:[1] they exhibit substantial inter-speaker

---

[1] These transcriptions are ad-hoc, as are many other transcriptions for the fricative vowels, because there is no consistently agreed-upon transcription for most of the fricative vowels. In this thesis, where no clear transcriptional default exists, I follow a convention of transcribing fricative vowels using a vowel symbol that suggests the fricative vowel's formant quality followed by a fricative symbol that suggests its fricative quality,

variation in the quality and intensity of frication, suggested by prior work on Sūzhōu Chinese to correspond to the anteriority of a supralaryngeal constriction made with either the tongue blade or the tongue dorsum (Wang, 1987; Ye, 1988; Qian, 1992; Hu, 2007; Ling, 2009). More anterior fricative vowels are generally reported as having a [ʑ] or [z]-like quality quite similar to a voiced fricative; more posterior fricative vowels appear to use what is essentially a modified [i] posture that generates a dorso-postalveolar constriction (Ling, 2009). Pilot data presented in the following chapter confirm this arrangement and additionally suggest that more anterior and laminal articulations may be innovative, while more posterior and dorsal articulations are conservative.

Articulatory resemblance of these postalveolar fricative vowels to the Sūzhōu Chinese fricative /ɕ/, instead of a tongue posture intermediate between a /i/ and a fricative, strongly suggests the "hacking" discussed in the previous section. That is, the articulatory strategy for a given Sūzhōu Chinese speaker's fricative vowels resembles much more closely their articulation of a fricative consonant also present in the speaker's inventory (usually /ɕ/). This "copying" of the motor program for lingual posture of some fricative consonant can be attributed to biases and pressures toward similarity of motor plans of two very similar segments, of the sort discussed in the preceding chapters. Over successive generations of acquisition, a bias toward "hacking" during L1 learning may even increase the incidence of fricative vowel variants that reuse tongue postures typical of /ɕ/.

To investigate the possibility of uniformity among these segments and change in the typical individual's degree of uniformity over time, in this chapter I provide background both on Sūzhōu Chinese and the vowels at issue as a prelude to an ultrasound experiment in the next chapter. After a brief discussion of the sociolinguistic context of Sūzhōu Chinese (Section 3.1), I give an overview of its tonal and segmental phonology (3.2). Particular attention is given to phonemicization of the vowels, especially the postalveolar fricative vowels and the similar apico-alveolar fricative vowels. Section 3.3 is given over to a discussion of fricative vowels in the various other languages in which they have been studied, including a discussion of their place in an expanded vowel space more generally, inclusive of fricative noise as an additional dimension of contrast. Finally, I consider in some detail the current understanding of the variation and historical development of the postalveolar fricative vowels in Sūzhōu Chinese (Section 3.4). An ultrasound experiment taking this linguistic situation as its backdrop, and informed in its hypotheses by the current understanding of variation in the Sūzhōu Chinese fricative vowels, is presented in the following chapter.

3.1 Sūzhōu Chinese

Both the milieu of Sūzhōu Chinese and its phonetics and phonology are relevant to the research questions at hand. Both are also fairly complex and are likely unfamiliar to the reader. In the first section of this chapter, I provide an overview of important linguistic and

---

joined with a tie bar.

contextual information on Sūzhōu Chinese. In order to facilitate access to relevant Chinese-language literature, I provide transliteration of relevant linguistic terms in Hànyǔ Pīnyīn 汉语拼音 romanization and simplified Chinese characters (hànzì 汉字) whenever possible. If referenced material was published in Chinese, its authors' names, its title, and its publication information are provided in hànzì in the bibliography.

3.1.1 Classification and location

Sūzhōu Chinese, commonly referred to as the "Sūzhōu dialect", is a language variety belonging to the Wú Chinese dialect family. Wú 吴 is a generally accepted subfamily of the Chinese languages within the Sino-Tibetan family (Simmons, 1999). The Wú dialects are spoken in most of Zhèjiāng 浙江 province, the southern half of Jiāngsū 江苏 province, and the province-level municipality of Shànghǎi 上海, as well as small parts of neighboring provinces (Figure 3.1) (Norman, 1988, p. 199; You, 2015). More specifically, it is a member of the Tàihú 太湖 dialect group, which is spoken around the lake of the same name, north to the Yangtze river, and east to the ocean and the city of Shànghǎi (Qian, 1992, pp. 2–3; You, 2015).

The term "Sūzhōu Chinese" is used here to refer to the variety of Tàihú Wú Chinese spoken in the urban core of Sūzhōu proper, to the exclusion of other Tàihú Wú varieties spoken elsewhere in Sūzhōu city (Sūzhōu shì 苏州市). Translation of the term "city" (shì 市) into English can result in ambiguity: there is no one-to-one correspondence between the term "city" as typically used in English and the Chinese administrative unit usually translated into English as "city." Chinese "cities" typically incorporate a core urban area, several smaller satellite urban areas or "sub-cities", and the surrounding hinterlands as a single unit. As such, the city of Sūzhōu actually includes five urban areas that an English speaker would call "cities": the primary urban area of Sūzhōu proper, in addition to four sub-cities (Zhāngjiāgǎng shì 张家港市, Chángshú shì 常熟市, Kūnshān shì 昆山市, and Tàicāng shì 太仓市). Inhabitants of the sub-cities speak Tàihú Wú dialects that are mutually intelligible with the dialect spoken in the old core of Sūzhōu but noticeably different from it in phonetic structure; it is thus critical that participants in the experiment in Chapter 4 have approximately the same residential history.

Even within the five districts that comprise the urban area of Sūzhōu itself, there is substantial lexical and phonological variation (Ye, 1988, pp. 18–67); participants were carefully recruited from Gūsū district (Gūsū qū 姑苏区), which contains the historical old city, to ensure a relatively homogeneous study population (see Section 3.4.1). The four other districts (Xiāngchéng qū 相成区, Hǔqiū qū 虎丘区, Wúzhōng qū 吴中区, and Wújiāng qū 吴江区) encompass secondary cores, outlying business districts and industrial areas, urban sprawl, and (in the case of Wújiāng) even farmland. Participants frequently contrasted the speech patterns typical of Gūsū district with those typical of the other districts, particularly Wújiāng.

The terms "Sūzhōu dialect" and "Sūzhōu Chinese" will be used interchangeably in this thesis, as they are in the resources cited. Usage of the term "dialect" should not be taken to

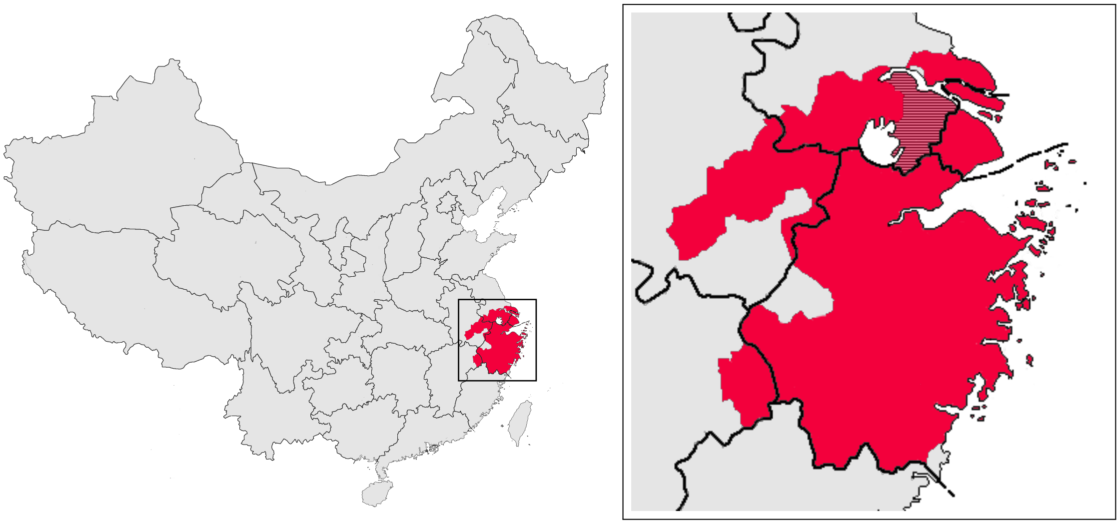


Figure 3.1: The approximate area within China where Wú dialects predominate (left), and the location of Sūzhōu city within the Wú area (right, in dark gray) (Yan, 1988; Zhao, 2008). Map derived with alterations from `https://commons.wikimedia.org/wiki/File:China_County-level.png`, by Wikimedia Commons user ASDFGHJ, under the image's CC BY-SA 3.0 license (see `https://creativecommons.org/licenses/by-sa/3.0/legalcode`).

imply that Sūzhōu Chinese and Standard Chinese (also known as Standard Mandarin) are mutually intelligible variants of the same language. Rather, the term "dialect" reflects the unique Chinese linguistic situation. While the Wú dialect group has a long history of phonological and lexical changes distinct from those undergone by the northern Mandarin varieties that form the basis of Standard Chinese, both Sūzhōu Chinese and Standard Chinese share hànzì 汉字 as a common writing system and cultural core. The traditional Chinese point of view foregrounds this deep, shared cultural heritage, with a literary standard language at its center, rather than the internal diversification of the many varieties of Chinese (see Norman, 1988, pp. 1–6).

3.1.2 Sociolinguistic situation and domains of use

Sūzhōu Chinese is the L1 of a majority of the inhabitants of Sūzhōu, but it is encroached on by Standard Chinese in a growing number of language-use domains. The growth of Standard Chinese usage has affected the domains of usage in which Sūzhōu Chinese is used, as well as the specific lexical forms employed by Sūzhōu Chinese speakers, with implications for the choice and suitability of experimental production materials in the next chapter.

L1 speakers of Sūzhōu Chinese typically have good command of spoken and written Standard Chinese as an L2, which is universally introduced as part of the primary school curriculum in the People's Republic of China (see Dwyer, 1998, p. 81). Many younger speakers, however, report that they acquired Standard Chinese in the home, simultaneously with Sūzhōu Chinese, from their Sūzhōu Chinese-native parents before their schooling began. Sūzhōu Chinese exhibits a modest amount of lexical influence from Standard Chinese (see Section 3.4.2). Further lexical influence, if not linguistic structural influence, likely lies ahead, as has already occurred for other Wú dialects in more intense contact with Standard Chinese (see Simmons, 1999 on Shànghǎi). Structural influence has likely already had small effects on some aspects of Sūzhōu Chinese, primarily at the level of lexical choice; a very small number of speakers appear to exhibit phonotactic patterns influenced by Standard Chinese phonotactics.

Although overlap in domains of use is increasing, the domains of use for Standard Mandarin and Sūzhōu Chinese still differ substantially, with the latter primarily used as the spoken language of day-to-day life and the former used for all official functions, including education, and nearly all writing. Chinese characters can be used to write both Sūzhōu dialect and Standard Chinese (Norman, 1988, pp. 75–77), but in practice most writing in the public sphere is “fixed and codified” to Standard Chinese grammatical norms (Dwyer, 1998, p. 72). Speakers of Sūzhōu Chinese are often uncomfortable with assigning a Sūzhōu Chinese “dialect reading” to characters due to the extremely low frequency with which these readings are situationally appropriate, Standard Chinese readings being the default. Some language-use domains relating to cultural heritage are still given over to the Sūzhōu dialect, particularly those that emphasize its historical prestige and internal artistic development. In particular, Sūzhōu táncí 苏州弹词, commonly known as píngtán 评弹, is a genre of musical storytelling that developed in Sūzhōu and is is spoken and sung in Sūzhōu Chinese; it remains popular both within and outside the city to the present day (Zhou, 2000; Bender, 2003).

The number of Sūzhōu Chinese speakers is not precisely defined, leaving its relative level of endangerment unclear. Available demographic data generally fall prey to the fallacy that all inhabitants of an area that traditionally speaks a given Chinese dialect are speakers of that dialect (see Yan, 1988). Given China’s recent history of internal migration, and my own anecdotal experiences in the city of Sūzhōu, this is almost certainly not the case: Sūzhōu, like other affluent coastal cities, hosts a large population of internal migrants from other dialect regions, and even a large number of native speakers of Standard Chinese.

The number of Wú Chinese speakers in the urban core districts of Sūzhōu city can be approximated from Ethnologue’s 2013 estimate of 69.7 million speakers of all Wú dialects (Simons and Fennig, 2015). Compared with the total population of traditionally Wú-speaking areas, as defined by Yan (1988), Yan (1981), and Zhao (2008), of 107.8 million[2], this gives a proportion of about 65%. If we extend this proportion to the population of the urban districts of Sūzhōu, we arrive at a figure of approximately 3 million speakers out of 4.6 million inhabitants. While this is admittedly a crude estimate, it is comparable to estimates provided by well-informed residents—about two million—and in line with historical estimates

---

[2] This figure is estimated using county-level population statistics from several provinces’ tabulations of the 2010 census of the People’s Republic of China (Shanghai Bureau of Statistics, 2012; Census Office of Zhejiang Province, 2012; Census Office of Jiangsu Province, 2012; Jiangxi Bureau of Statistics, 2012).

Upper register:			
Yīnpíng	阴平	疤 [po]^{44}	‘scar’
Yīnshǎng	阴上	靶 [po]^{51}	‘target’
Yīnqù	阴去	坝 [po]^{523}	‘dam’
Yīnrù	阴入	八 [poʔ]^{5}	‘eight’
Lower register:			
Yángpíng	阳平	爬 [bo]^{23}	‘crawl’
Yángqù	阳去	罢 [bo]^{231}	‘stop’
Yángrù	阳入	薄 [boʔ]^{3}	‘despise’

Figure 3.2: Relative pitch level contours (5 = highest, 1 = lowest) for the seven citation tones of Sūzhōu Chinese. Pitch contour values are modified, based on my own observations, from Wang (1987), Ye (1988), and Qian (1992). Example lexemes from Ye (1988).

by experts, assuming population growth (Zhengzhang, 1988).

3.2 Sūzhōu Chinese phonology

Below, I briefly discuss in turn the tonal phonology of Sūzhōu Chinese (Section 3.2.1) and the segmental phonology of its consonants (Section 3.2.2) and vowels (Section 3.2.3). All three aspects of Sūzhōu Chinese phonology have some bearing on the issue of the articulation and acoustics of the fricative vowels, which are discussed briefly in Section 3.2.4 and elaborated upon in Section 3.3.

3.2.1 Lexical tone and register

Sūzhōu Chinese is a lexical tone language with a register contrast dividing the tones into two sets, generally called yīnshēng 阴声 and yángshēng 阳声 or the upper and lower registers, respectively. The seven contrastive tone contours on syllables spoken in isolation are given in Figure 3.2, divided by register. In addition to their differences in pitch level and contour, the distinction between the yīnshēng and yángshēng registers is thought to be cued by the presence of aperiodicity in the signal. Lower register productions in the various Wú Chinese dialects exhibit breathiness or murmur (Cao and Maddieson, 1992; Chen, 2014, pp. 116–17; Chen and Gussenhoven, 2015, pp. 322–23, 333–34) or other non-modal phonation (Rose, 1989). Although instrumental studies on phonation in the Sūzhōu Chinese registers are lacking, Wang (1987) likewise describes the Sūzhōu Chinese upper register as articulated with “tightened phonation” and the lower register as having “turbidity” or breathy phonation, the contrast being particularly clear on syllables with semivowel or null initial consonants (pp. 43–44).

	Bilabial	Labiodental	Alveolar	(Alveolo-)palatal	Velar	Glottal
Plosive	p pʰ b		t tʰ d		k kʰ g	(ʔ)
Nasal	m		n	ȵ		
Affricate			ts tsʰ	tɕ tɕʰ dʑ		
Fricative		f v	s z	ɕ		h (ɦ)
Lateral			l			
Semivowel				j ɥ	w	

Figure 3.3: Sūzhōu Chinese consonantal phonemes, modified from Ye (1988).

The lower register is (in part) signaled by the presence of breathy phonation throughout Wú Chinese. Based on current research, Sūzhōu Chinese's register contrast can be described as cued solely by phonation on the syllable nucleus, which is not the situation in the more thoroughly studied Shànghǎi dialect (see Section 3.2.2): in the latter, voiced obstruents that co-occur with the lower register are consistently voiced in at least some environments (Sherard, 1972, p. 78; Chen and Gussenhoven, 2015). The most typically described arrangement for Wú generally, Sūzhōu included, is devoicing of all "voiced" stops when these occur as the first segment in a phonological phrase (Wang, 1987, p. 43; Ye, 1988, p. 106; Chen and Gussenhoven, 2015, p. 324). Since this adds some aperiodic noise to the signal during vowel production, register is significant for later discussion of the Sūzhōu Chinese fricative vowels, since both the lower register and the fricative vowels may be cued by the presence of different types of aperiodic energy that may nonetheless similarly impact whole-spectrum measures of aperiodicity.

3.2.2 Onset consonants

Sūzhōu Chinese has 28 consonant phonemes which may occur in the syllable onset (Figure 3.3). Plosives and affricates at all places exhibit a three-way contrast between aspirated voiceless, unaspirated voiceless, and voiced phonation, with the exception of unattested */dz/ and */ʑ/. Voiceless aspirated and unaspirated obstruents occur as onsets to upper-register or yīnshēng syllables, whereas the voiced obstruents occur as onsets to lower-register or yángshēng syllables, as discussed above. Non-obstruents (nasals, the lateral approximant, and the semivowels) may occur in syllables of either register. Sūzhōu Chinese's voiced obstruents /b d g dʑ v z/ are in fact typically described as voiceless (Wang, 1987, p. 43; Ye, 1988, p. 106) when they occur at the beginning of a phonological phrase, which agrees with my impressions based on the data recorded for this experiment.

The phonemic status of the glottal onsets [ʔ] and [ɦ] has been a source of disagreement in previous analyses of Sūzhōu Chinese, as they can be attributed to the register with which they co-occur. Two different analyses exist for the phonological status of [ʔ] and [ɦ]. In the first, [ʔ] and [ɦ] are phonemic onsets /ʔ/ and /ɦ/ that occur in upper and lower register syllables, respectively (Ye, 1988; Qian, 1992). In the second analysis, [ʔ] and [ɦ] both occur as phonetic onsets to phonologically onsetless syllables, with the contrast between the apparent onsets actually due entirely to the registers' different laryngeal settings (Wang, 1987; Xing, 2014). I prefer the latter of the two analyses and employ it here. The third glottal onset, /h/, is unproblematic since it consistently occurs in the upper register and contrasts with upper-register /ʔ/ or upper-register zero onset, depending on the analysis employed.

Coronal obstruents are divided into two contrastive series, the alveolars and alveolopalatals. A sporadic palatalization sound change that is not, to my knowledge, reported in the previous literature merges alveolar fricatives and affricates with the equivalent alveolopalatal segments before high front vowels /i y/ and the fricative vowel /i͡ʑ/ (no sequences of the alveolars and /y͡ʑ/, such as *[sy͡ʑ], are attested). Speakers vary considerably in their application of this change, with younger speakers exhibiting the most alveolopalatal productions. Some older participants retain a contrast between the alveolars and alveolopalatals before /i͡ʑ/ and /i/ (i.e. /si͡ʑ/ versus /ɕi͡ʑ/; /si/ versus /ɕi/; most speakers do not, although it is relatively common for speakers to palatalize before /i͡ʑ/ but not /i/.

3.2.3 Vowels

Existing descriptions of Sūzhōu Chinese phonology adopt the analytical stance of the Chinese phonological tradition that the syllable onset and rhyme are the basic units of analysis, rather than contrastive consonantal or vocalic phonemes. I choose instead to reanalyze multisegmental rhymes as sequences of phonemically contrastive vowels and coda consonants. Under this analysis, Sūzhōu Chinese has 15 monophthong vowel phonemes, several of which are of interest for the experiment in the following chapter, and two unit diphthong phonemes (Figure 3.1). The phoneme /ø/ is produced more central than the cardinal vowel suggested by its transcription (Ling, 2009, pp. 28–29); closer to the central vowel [ɵ], as used in Qian (1992). The phoneme /a/, when in open syllables, is produced somewhat fronter and higher than its transcription would suggest, often sounding similar to the usual phonetic value of standard American English /æ/.

Four of the monophthongs are "checked," or shortened and followed by a glottal stop phrase-finally; syllables of this type only co-occur with the rùshēng 入声 tones (see Section 3.2.1). The general Wú Chinese tendency to elide the glottal stop when it does not occur before pause often leaves the "checked" vowels as merely shortened relative to "unchecked" vowels when followed by another syllable (Chao, 1967). Furthermore, the checked vowels in the closely related Shànghǎi dialect do not appear to be reduced allophones of the open-syllable vowels: according to the acoustic analysis in Chen (2008), their characteristic formant frequency values cannot be attributed to simple reduction due to shortening. Given this evidence, and given a lack of evidence for allophony with the non-checked vowels, I tentatively treat the Sūzhōu Chinese checked vowels as distinct phonemes with their own targets for duration, glottalization, and perhaps formant values.

Two monophthongal vowels (one phonemic) are the so-called APICAL VOWELS and two

		Unrounded	Rounded		
Coronal	Anterior	ɿ	[ʮ]		
	Posterior	i͡ʑ	y͡ʑ		
		Front		Central	Back
		Unrounded	Rounded		
Dorsal	High	i	y		
	Mid	ɛ	ø	ə əʔ	o oʔ
	Low	a aʔ			ɑ ɑʔ
Other: Diphthongs əu, eɪ ∼ øʏ; Labial [ə͡v], [ə͡β]					

Table 3.1: The vowel phonemes of Sūzhōu Chinese. Provisionally, the four vowels that occur in rùshēng (checked) rhymes are treated as separate phonemes, as in Chen (2008)'s analysis of Shànghǎi dialect vowels. Vowels that are not contrastive on distributional grounds are given in square brackets [ ].

plus an allophone are so-called FRICATIVE VOWELS[3], all of which are produced with some supralaryngeal constriction. All apical and fricative vowels have a muted, lowered-intensity quality to their periodic spectral components, and also exhibit fricative noise originating at the appropriate place of articulation is most frequently produced. The "fricative" postalveolar vowels and apico-alveolar vowels are described in greater detail in Sections 3.3–3.4.

The unrounded and rounded apical vowels, transcribed in this thesis using the traditional Chinese phonetic symbols [ɿ] and [ʮ], have an apico-alveolar constriction similar to a /z/ and could be transcribed as syllabic rounded and unrounded alveolar fricatives with a loose degree of constriction, i.e. syllabic, lowered [z], [z^w]; both exhibit noticeable strident frication with a [z]-like quality. The postalveolar "fricative vowels" have a more posterior constriction made with the tongue blade or dorsum, giving an auditory quality between somewhat opened syllabic [ʑ] and [ʝ] for the unrounded vowel and slightly opened syllabic [ʑʷ] and [ʝʷ] for the rounded vowel. These are transcribed throughout as /i͡ʑ/ and /y͡ʑ/, respectively, for lack of a convenient IPA symbol. The labial "fricative vowel" has labiodental constriction or bilabial compression, matching the place of its onset; the bilabial variant is sporadically produced with some bilabial trilling, i.e. as syllabic [ʙ] (Ling, 2009, p. 60). I transcribe these as [ə͡v] and [ə͡β], respectively, to reflect the labial constriction combined with a neutral tongue position.

There are strong distributional grounds for analyzing the rounded "apical" vowel [ʮ] as an allophone of the rounded "fricative" vowel /y͡ʑ/ that occurs only after alveolar fricative and affricate onsets (see Section 3.2.4). Similarly, the labiodental fricative vowel [ə͡v] and bilabial fricative vowel [ə͡β] are in complementary distribution with [əu]—the allophone [ə͡v] occurs

---

[3] Note that the term "fricative vowel" is used in a more restrictive sense than the usage I adopt starting in Section 3.3.

after labiodental fricatives, [ə͡ʙ] after bilabial stops, and another allophone [əu] elsewhere—leading me to treat [ə͡v] and [ə͡ʙ] as allophones of /əu/. This analysis is at odds with the usual presentation of the rhymes as contrastive in formal descriptions of Sūzhōu Chinese (e.g. Wang, 1987; Qian, 1992).

3.2.4 Phonotactics and phonemic analysis

Due to phonotactic restrictions on onset-rhyme co-occurrence and the co-occurrence of segments within rhymes, the number of possible syllables in Sūzhōu Chinese is far lower than implied by the number of onsets and rhymes presented in Figures 3.3–3.1. All Sūzhōu Chinese onsets are simple (i.e. C-); rhymes have the maximal shape -GV{ʔ|N}, that is, a monophthong (with or without a coda glottal stop; nasal or oral) that may preceded by a semivowel in various combinations. Only the phonemes /ɛ, ə, a, ɑ, o, ø/ may be preceded by a glide, and only the more restricted set /ə, a, ɑ, o/ occurs in nuclei that are "checked" (with a glottal coda) or nasalized.

Both glottalization and nasality are debatably properties of the preceding vowel rather than coda consonants (see Chen (2008) on Shànghǎi Chinese). The coda "glottal stop" has been analyzed above as a specification of a subset of the vowels. Nasality's place of articulation (or lack thereof) is predictable from the vowel itself: following /o/, nasality takes the form of a velar nasal stop [ŋ]; following /ə/, it takes the form of an alveolar nasal stop [n]; and the low vowels /a ɑ/ are themselves nasalized, with no following nasal stop (Wang, 1987, p. 41–2; Ye, 1988, p. 106–7; Qian, 1992, p. 37–38).

All consonants shown in Figure 3.3 (Section 3.2.2) are attested in onset position; however, co-occurrence restrictions with rhymes limit the number of syllables actually attested. The most significant restriction for purposes of this thesis prevents co-occurrence of alveolar fricative/affricate onsets and velar onsets with the high front vowels /i, y/, complex rhymes starting with high front semivowels /j, ɥ/, and the "fricative vowels" /i͡z̞, y͡z̞/ (summarized in Figure 3.2). Alveolopalatal initials only occur with the monophthongs /i͡z̞, y͡z̞, i, y/ and complex rhymes with medials -j- and -ɥ- (e.g. -joŋ, -ɥəʔ, etc.). The velar initials and /h/ do not occur before this same set of rhymes.

While alveolar fricatives and affricates generally contrast with the alveolopalatals before the high front vowels /i, y/ and the semivowels, many speakers also palatalize these to alveolopalatals, setting the stage for a complementary distribution of three onset types: alveolar fricatives and affricates, along with velars and /h/, are in complementary distribution with the alveolopalatal affricates and fricatives, with the latter occurring before the high front vowels and fricative vowels and the others occurring elsewhere. Additionally, alveolar fricative and affricate initials /ts, ts^h, s, z/ are the only initials which co-occur with the "apical" vowels /ɿ/ and [ʮ], which for these same speakers forms an additional complementary distribution of the "apical" and "fricative" vowels.

The "apical vowels" could thus each be analyzed as allophones of the "fricative vowel" with the same labial activity; that is, for some speakers, /i͡z̞/ having the allophone [ɿ] following alveolar fricatives and affricates and [i͡z̞] elsewhere, and an analogous mapping of /y͡z̞/ to [ʮ]

	/i/, /y/	/i͡ʑ/, /y͡ʑ/	/ɿ/, [ʮ]	/jV/, e.g. /jəʔ/, /jã/	Other /V/, e.g. /əʔ/, /ã/
Labials	Y	Y	N	Y	Y
Alv. stops, /n/, /l/	Y	Y	N	Y	Y
Alv. fricatives and affricates	Y ∼ [ɕ]	Y ∼ [ɕ]	Y	Y ∼ [ɕ]	Y
Alveopal. fricatives and affricates	Y	Y	N	Y	N
Velars, /h/	N	N	N	N	Y
No initial	Y	Y	N	Y	Y

Table 3.2: Distribution of onsets by place following high front vowels, high front semivowels, and "fricative vowels". Dental fricatives and affricates that are variably produced as alveolopalatal are indicated with ∼ [ɕ].

and [y͡ʑ]. I have adopted this analysis for /y͡ʑ/ above in Section 3.2.3, but do not extend this to /i͡ʑ/ and /ɿ/ given that many speakers still have minimal pairs contrasting /i͡ʑ/ and /ɿ/, such as 丝 [sɿ^{44}] "thread" versus 西 [si͡ʑ^{44}] "west". Where /y͡ʑ/ occurs only with zero onsets and alveolopalatal onsets, /i͡ʑ/ may occur after zero onsets, bilabials and labiodentals /p p^h b m f v/, and alveolopalatals /tɕ tɕ^h dʑ ɕ ɲ/; many speakers additionally allow the alveolar stops /t t^h d/ to precede /i͡ʑ/. The relatively free distribution of /i͡ʑ/ makes it possible to test hypotheses relating to articulatory reuse, as will be discussed in the next chapter.

Based on complementary distribution alone, the alveolopalatals /tɕ tɕ^h dʑ ɕ ɲ/ could also be analyzed as palatalized allophones of /k k^h ɡ h ŋ/, respectively. However, I opt to retain alveolopalatals as a separate phonemic series to simplify the assumptions being made about speakers' mental representations of their speech production activity. In Wú Chinese, Sūzhōu Chinese included, /tɕ tɕ^h dʑ ɕ ɲ/ develop not only from the equivalent velars, but also from alveolar *ts, *ts^h, *z, *s, *n, which frequently palatalize before high front vowels (Ballard, 1969). The alveolopalatal series is thus created through the merger of two other series in a palatalizing context. Sūzhōu Chinese speakers cannot extract information on whether a given alveolopalatal is underlyingly alveolar or velar, given that a static distributional restriction obtains, and there are no morphophonological alternations upon which to base identification of alveolopalatals with some other series of consonants.

This decision is not without precedent, since both the phonotactic restrictions on velars and coronal fricatives and affricates and the resulting analytical dilemma are not uncommon in the Chinese dialects (Chao, 1934; Norman, 1988; Ao, 1993; Duanmu, 2007; Chen and Gussenhoven, 2015). In particular, Ao (1993) discusses at length the problem of non-unique phonemic solutions such as the one presented by the distribution of velars in many dialects of Chinese, including Sūzhōu Chinese: there is no principled way to determine whether alveolopalatals are underlyingly velar or alveolar.

Syllabic /l/ or a sequence /əl/ is claimed to occur in a handful of morphemes by most existing analyses (Wang, 1987, p. 41; Ye, 1988, p. 140; Ling, 2009, pp. 9–10), for example 二 [əl^{213}] 'two'. Occasionally this rhyme is implied to be rhotic rather than lateral (Qian, 1992, p. 37; Xing, 2014, p. 7). However, none of the speakers recorded for the following chapter's experiment employed a lateral approximant in their readings of the character 儿 'child', given as [əl^{23}] in Xing (2014). A handful of speakers did produce a Standard Chinese reading [ɑɻ^{23}] or a similar reading with adaptations to Sūzhōu phonology (which lacks rhotics) as [ɑʊ^{23}]. Most speakers instead provided the typical Sūzhōu Chinese reading of [n̩ɪ͡ʑ^{23}].

3.3 Phonetic characteristics of fricative vowels

A notable feature of Sūzhōu Chinese phonology is the large number of segments that occur in phonotactic positions typical of vowels but give the auditory impression of voiced syllabic central approximants or voiced syllabic strident fricatives. Below, I refer to these segments as a group as fricative vowels, for their supralaryngeal constrictions and apparent fricative noise targets. I locate these segments in a reworked and expanded model of the vowel space based around a full possibility space of constriction locations, rather than the prevalent model of exclusively dorsal constrictions. A large majority of these possible vowels have non-dorsal constrictions, having major constrictions produced instead by manipulation of some other articulator such as the tongue tip, lips, or pharynx.

In this section, after a review of existing research into the phonetic and phonological properties of fricative vowels, including the so-called "apical vowels" (Section 3.3.1), I address the role of fricative noise as a production goal for vowels (Section 3.3.2) and propose a novel vowel classification scheme taking possible frication targets and place of articulation into account (Section 3.3.3). Finally, in Section 3.3.4, I provide an overview of the attested types of fricative vowels, their distribution in the world's languages, and their position within the classification laid out in Section 3.3.3. I discuss each major place of articulation in turn along with its typical active articulator: apico-alveolar vowels, postalveolar vowels, and labial vowels, inclusive of labiodental vowels and bilabially compressed vowels. A synchronic and diachronic description of the fricative vowels in Sūzhōu Chinese specifically, building off of this general foundation, follows in Section 3.4.

3.3.1 Prior research on fricative vowels

A variety of fricative vowels are common in the Chinese dialects, and as such, the nomenclature for these sounds has been shaped by research in the Chinese phonological tradition. They are conventionally divided into two categories: the relatively common APICAL VOWELS or shéjiān yuányīn 舌尖元音 and the less common FRICATIVE VOWELS[4] or mócāhuà

---

[4] Again, note that I expand this term's scope in later sections to refer to all vowels with fricative noise targets; this includes most "apical" vowels and even some dorsal vowels.

yuányīn 摩擦化元音. The apical vowels are by far the more commonly discussed and attested of the two groups, primarily owing to their occurrence in Standard Chinese (and a majority of Chinese dialects, as detailed below). Previous research on Standard Chinese apical vowels has found that they are generally produced with a raised tongue tip, similar to a central apico-alveolar or apico-postalveolar approximant but with some differences in phonetic implementation. Apical vowels tend to be produced with some amount of fricative noise resembling the spectrum of [z] or [ʐ], and the lingual posture has been reported to be most similar to a strident fricative (Zhou and Wu, 1963; Lee-Kim, 2014; Faytak and Lin, 2015) rather than a rhotic or lateral approximant.

Apical vowels characteristically have a reduced phonotactic distribution in varieties of Chinese, occurring only as allophones of a high vowel (most frequently /i/) following a language's apico-dental, apico-alveolar, or apico-postalveolar fricatives and affricates (i.e. the series /s ts tsʰ/, /ʂ tʂ tʂʰ/). At most two types of "apical" vowel are known to contrast with each other in any one language, covarying with the place of the syllable onset: apico-dental or apico-alveolar [ɿ] after apico-dental or apico-alveolar /s ts tsʰ/, and apico-postalveolar [ʅ] after apico-postalveolar /ʂ tʂ tʂʰ/, respectively.[5] Rounded versions at both places of articulation occur less frequently: rounded apico-dental or apico-alveolar [ʮ], and rounded apico-postalveolar [ʯ]. The fact that these vowels are homorganic with a preceding consonant suggests that they owe their development to coarticulation of high vowels with immediately preceding consonantal fricatives (Chen, 1976; Yu, 1999) rather than the development of a production target for fricative noise.

How to phonologically describe the manner of articulation of apical vowels has been a topic of some debate, in part because there is no consensus on how to describe their essential acoustic phonetic features. The view from Chinese linguistics is that apical vowels and fricative vowels are both quintessentially vowels (as the terms in use would imply), largely on the basis of phonotactics: they appear in vowel-like positions and host lexical tone contrasts. This view is encountered in most phonetics research carried out in China for a Chinese audience (Zhou and Wu, 1963; Zhao, 2007; Zhu, 2015) as well as in more functionally oriented research published elsewhere (Karlgren, 1926; Howie, 1976; Wu, 1994). The traditional symbols [ɿ, ʅ, ʮ, ʯ] used in these lines of research imply vowel status, and are a standard, accepted part of the phonetic alphabet typically used in phonetic-phonological description in China, inspired by the Swedish dialectological symbols originally employed by Karlgren (1926) in his description of Standard Chinese, but which are not official IPA symbols (Pullum and Ladusaw, 1996).

In research more oriented toward generative linguistics and formal phonology, on the other hand, the most common analysis of apical vowels is as syllabic consonants, typically voiced syllabic fricatives, e.g. [z], [ʐ] (Chao, 1934; Pulleyblank, 1984; Wiese, 1997; Duanmu, 2007). This view usually assumes featural spreading and "annexation" of the syllable nucleus

---

[5] The apico-postalveolar vowels, fricatives, and affricates are traditionally described as "retroflex" in Chinese linguistics, but are more generally flat apico-postalveolar, in contrast to the "true" retroflexes found in, e.g., the Dravidian languages (Ladefoged and Maddieson, 1996).

by the co-occurring fricative or affricate onset, such that the nucleus can be described as an underlyingly underspecified or "zero" syllabic (Li, 1966; Pulleyblank, 1984; Wiese, 1997), and phonetically as a continuation of the onset (see Lee-Kim, 2014, p. 263).

A third way has emerged from recent articulatory evidence: where apical vowels are described as syllabic central approximants. Lee-Kim (2014) argues for an approximant transcription and representation for the Standard Chinese apical vowels. Her ultrasound data confirm that apical vowels exhibit essentially the same tongue shape as the fricative consonants they follow. Based on auditory impression and visual impression of spectrograms obtained from recorded data, Lee-Kim argues that the apical vowels typically lack fricative noise and opts for the approximant analysis to reconcile the acoustic and articulatory data, supposing a minute loosening of constriction degree as a speaker begins to produce the apical vowel.

In contrast to apical vowels, fricative vowels are defined as exhibiting substantial turbulent airflow and fricative noise in comparison to "conventional" (dorso-palatal, etc.) vowels. Fricative vowels are typically taken to include both vowels produced with a postalveolar constriction and some produced with a bilabial or labiodental constriction. Postalveolar, bilabial, and labiodental articulator configurations all typically generate fricative noise at the constriction location, and vowels with bilabial compression may exhibit bilabial trilling if aerodynamic conditions permit. Like apical vowels, the postalveolar vowels may be unrounded or (less commonly) rounded. Unlike apical vowels, fricative vowels are typically not allophones of another vowel, and display relatively unrestricted phonotactics; they may follow a varied set of onset consonants, such that their development cannot be pinned on assimilation to a preceding fricative or affricate consonant of a given place of articulation.

Nomenclature for the "fricative vowels" varies depending on the researcher and the linguistic area at issue. Most typically, the term mócāhuà yuányīn 摩擦化元音 (roughly "frictionized vowel") is employed in the Chinese-language literature. Zhu (2004) employs the term gāodǐng chūwèi yuányīn 高頂出位元音, roughly "off-the-chart vowels", in reference to the traditional vowel space trapezoid that can be produced using an unimpeded central channel defined by F1 and F2 frequency values. The term dài cā yuányīn 带擦元音 "friction bearing vowels" is also occasionally used (Hou, 2009), and (Rose, 1982) uses the term "strident vowels" for the Zhènhǎi 镇海 variant of the Níngbō 宁波 Wú dialect. The term "fricative vowel" itself has also been applied to the same type of vowel in West African languages (Connell, 2007).

Where "apical" vowels have conventionalized symbols (e.g. [ɿ]), the transcription of fricative vowels varies widely. If there is not a contrastive [i] or [u] quality in a given language, researchers will often simply use the closest cardinal vowel to stand in for the less common postalveolar or labial vowel. This is an especially common representation in descriptions of Chinese dialects, particularly Wú, as well as in Swedish dialectology, where the *Viby-i* and *Göteborges-i* are considered essentially /i/-like and are as such transcribed as [i] (e.g., Björsten and Engstrand, 1999). Less commonly, a sequence of the closest cardinal vowel and a fricative that accords with auditory impression of the vowel is used instead: for instance, Qian (1992) transcribes postalveolar vowels in Wú Chinese with [i_z] or [i_j], and Connell

(2007) transcribes the postalveolar vowel of Len Mambila as [ʒ͡ɨ] and the labiodental vowel as [v͡ʉ]. Less common strategies include denoting the fricative vowel as a syllabic consonant (Dell, 1994) or inventing a phonetic symbol (Rose, 1982). Zhu (2015) inventories a wide variety of other transcriptional representations found in the Chinese literature.

3.3.2 The role of fricative noise targets

There is some confusion over whether the "apical vowels" of Standard Chinese, which are often described as impressionistically fricated, actually have fricative noise targets or not, in spite of their having essentially fricative-like tongue postures. Lee-Kim (2014) has described Standard Chinese "apical vowels" as exhibiting only fricative noise that can be attributed to carryover from the fricative or affricate onset. A vowel I have observed in my own fieldwork in Oku, a language of northwestern Cameroon, also does not exhibit fricative noise but gives approximately the same percept as the Standard Chinese apico-alveolar "apical vowel". Yet, substantial fricative noise clearly characterizes some other "apical vowels", such as those found in Sūzhōu Chinese (Hu, 2007; Ling, 2009) and the Héféi 合肥 dialect of Jianghuai Mandarin (Hou, 2009). Faytak and Lin (2015) even observe that some Standard Chinese speakers' apical vowels exhibit frication that cannot be attributed to carryover from the onset.

The sheer amount of research into the Standard Chinese apical vowels ignores the fact that because its apical vowels lack frication, Standard Chinese is likely an exception rather than the norm. The apical vowels in many other Chinese dialects exhibit more frication than in Standard Chinese (Rose, 1982; Ling, 2009; Hou, 2009), and fricative vowels, as the term implies, are noted for their aperiodic noisiness as a rule (see previous section). Whether fricative noise is required to set apart a given non-dorsal vowel (e.g. [ɿ]) from otherwise acoustically similar dorsal vowels (e.g. [i])—in other words, whether a given [ɿ] can be considered a fricative vowel or just a non-dorsal vowel—may largely depend on the functional load of the contrast between the non-dorsal vowel and the dorsal vowel (see Section 3.4). For instance, the apical vowels in Standard Chinese are completely predictable from context, and their contrast with non-apical high vowels bears essentially no functional load. However, the noisier Suzhou Chinese fricative vowels are relevant to a number of minimal and near-minimal contrasts (see Section 3.2). Given that the Standard Chinese data point does not seem to extend to other attested examples of non-dorsal vowels, I assume that fricative and apical vowels in general have fricative noise targets. (Other evidence from diachrony for this analysis is given in Section 3.4.2.)

Aerodynamic factors may also interfere with the production of audible fricative noise in fricative vowels even if a target for fricative noise production is present, further muddying the picture. Like most vowels, fricative vowels are typically modally voiced. It is well known that production of voicing and the generation of turbulence from a supralaryngeal constriction are antagonistic due to the double air pressure drop required along the vocal tract: first across the glottis for the production of voicing, and then across the supralaryngeal constriction for generation of turbulent flow (Catford, 1977; Ohala, 1983). The production of modal voice in fricative vowels may interfere with the goal of fricative noise production, as in voiced fricatives more generally. Failure to produce turbulent flow, or even less turbulent flow than dictated by the target, may give rise to the percept of a voiced approximant (Catford, 1977, p. 121–25). Vowels produced with bilabial constrictions may also produce turbulent airflow that is outside of the oral cavity altogether, leaving it unable to strike an obstacle and generate perceptually consequential noise (Catford, 1977, p. 119, Shadle, 1990).

3.3.3 A revised nomenclature for fricative vowels

The existing terms "apical vowel" and "fricative vowel" ostensibly categorize the fricative vowels by their active articulator. However, actual usage of the terms is rife with inconsistencies, particularly for lingual articulations. Many "apical vowels" are in fact apico-postalveolar, the Standard Chinese "retroflex" apical vowels being a clear example. Some "fricative vowels" are apico-alveolar and are even transcribed using the symbols traditionally reserved for "apical vowels," such as the "fricative vowel" [ɿ] in Héféi Chinese (Wu, 1995; Hou, 2009). In practice, the term "apical vowel" seems to be used to describe a specific combination of phonotactics and broad articulator class: the phonotactics of the "apical vowels" in Standard Chinese may have played an outsize role in determining this nomenclature for these vowels. Any vowel (apical or laminal) with a constriction produced by the tongue tip or blade that is obligatorily preceded by a homorganic fricative or affricate—precisely the distribution of the "apical" vowels in Standard Chinese—may be called an "apical vowel"; anything else distributed according to any other phonotactic rules tends to be called a "fricative vowel".

The use of "fricative vowel" as a catch-all category for non-"apical" vowels with fricative targets also gives short shrift to the attested range of supralaryngeal constrictions that may accompany vowels: postalveolar, bilabial (with vertical compression), labiodental, and possibly others. It arbitrarily cuts vowels with apical, most canonically apico-alveolar, constrictions out from this broader set, and labial vowels are rarely explicitly included in this category at all (although see Zhu, 2004), even though they often have distributional or phonotactic similarities to alveolar or postalveolar vowels with fricative noise targets and appear to arise diachronically in a similar fashion (see Section 3.4).

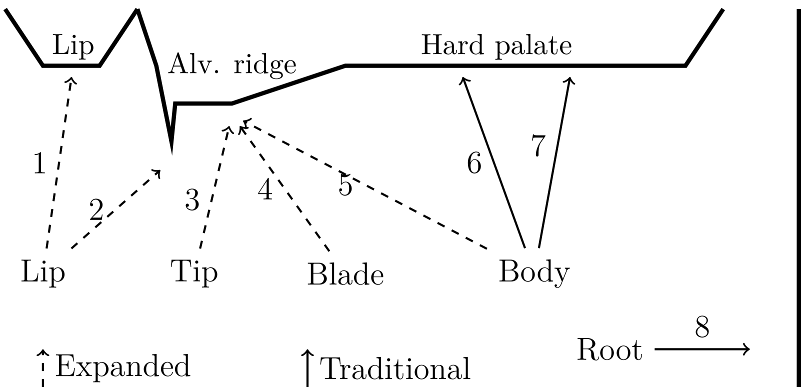
Furthermore, the terms "apical vowel" and "fricative vowel" are not part of a unified framework for possible vowel articulations, but rather an appendix to a framework dominated by the assumption that vowel quality is primarily determined by dorsal articulation, or the positioning of the highest point of the tongue body (e.g., Jones, 1956; Lindblom and Sundberg, 1971; Lindau, 1978, pp. 542–43; Honda, 1996). This framework ignores the fact that additional constrictions can readily be made using other parts of the vocal anatomy. For instance, the contributions of the tongue root and laryngeal musculature to the articulation of vowels are easily separable from the contributions of the dorsum (Esling, 2005). One could add that the other parts of the lingual articulator are completely undifferentiated in the dominant dorsal-only model, and actions of the lips beyond protruded rounding are only rarely commented on (Lindau, 1978; Linker, 1982).

To capture the full range of active and passive articulators actually used across the vowel space, I propose a revision to the division and characterization of the vowel space incorporating aspects of Esling (2005). Analogous to Zhu (2004)'s "off-the-chart vowels", I create an umbrella category for vowels outside of the canonical "dorsal" vowel space, or non-dorsal vowels. I name this category as such for their narrowest supralaryngeal constrictions being produced with active articulators other than the tongue dorsum, such as the tongue blade, tongue tip, or the lips. This category can also be extended to Esling (2005)'s "retracted" vowels, which I call pharyngeal, whose major constriction is produced using the pharynx and laryngeal articulator (i.e. in the range [ɜ]–[ɑ]). Non-dorsal vowels are accordingly divided into a range of places according to the articulator that is the major determinant of vowel quality: labial (with variants bilabial and labiodental), coronal (with variants on a continuum between alveolar and postalveolar, the specific active and passive articulators depending on the language), and pharyngeal (corresponding to Esling's "laryngeal" or "retracted"). This classification is laid out in Figure 3.4.

Vowels of various places of articulation may have (or lack) a fricative noise target. I define the set of vowels with fricative noise targets as fricative vowels. This converts the term from a catch-all for vowels having some fricative noise that are not "apical vowels" to a superordinate term that has clear subtypes organized by place. While it is theoretically possible for a back dorsal ("velar") or pharyngeal vowel to be constricted to the point where it produces frication, e.g. [ɯ] with [ɣ]-like fricative noise, this feature is not attested as a target for speech sounds to my knowledge, and even the development of frication in front dorsal vowels is unusual (though the Sūzhōu Chinese fricative vowels may be dorso-postalveolar for some speakers; see Section 3.4.1). Development of the more anterior or rostral places of articulation and a fricative noise target seem to go hand-in-hand. I do not attempt to transcribe this frication to avoid an overabundance of diacritics, and because it can be assumed for most non-dorsal vowels. It is also likely that intensity of fricative noise is a continuous acoustic dimension along which vowels may have targets, rather than a single parameter reducible to differences in articulation: some labial and coronal vowels may lean heavily on frication as a cue to category, while others may not.

I in turn divide the dorsal vowels, which make up most of the canonical vowel space, into three sets, following Esling (2005), and on the basis of known coordinative activity of external muscles affecting the placement of the bulk of the tongue dorsum. The tongue dorsum is pulled to the front for some vowels (e.g. in the range [i]–[æ]), primarily through action of the anterior genioglossus (Honda, 1996); the dorsum is pulled back (and raised) for some vowels, (e.g. in the range [u]–[o]), through the action of the styloglossus or intrinsic tongue muscles (Honda, 1996; Takano and Honda, 2007). A third set of vowels falls in between, likely owing to coordinated activity of both fronting and raising/backing activity (e.g. in the range [ɨ]–[ə]).

This approach to the vowel space is grounded in possible articulatory configurations for vowels. This way of thinking about the "fricative vowels" of Sūzhōu Chinese—as a series of postalveolar segments with the goal of producing some intensity of strident frication—is critical for examining articulatory reuse in series of sounds in phonological inventories. The



Key	Major articulators		V attested in
	Active	Passive	
1	Lower lip	Upper lip	Sūzhōu Chinese, Kejom
2	Lower lip	Teeth	Sūzhōu Chinese, Kejom, Kom
3	Tongue tip	Alveolar ridge (etc.)	Sūzhōu Chinese, Standard Chinese, Oku
4	Tongue blade	Alveolar ridge (etc.)	Sūzhōu Chinese (var.), Kom
5	Tongue body	Alveolar ridge (etc.)	Sūzhōu Chinese (var.)
6	Tongue body	Anterior palate	Numerous, e.g. [i]
7	Tongue body	Posterior palate	Numerous, e.g. [u]
8	Tongue root	Pharynx	Numerous, e.g. [ɑ]

Figure 3.4: Lines of articulatory action employed in the world's vowels, including both the traditional vowel space (solid arrows, 6–8) and the expansions to the vowel space discussed here (dashed arrows, 1–5), and attestations of these vowels that have been discussed above. The alveolar ridge as a passive articulator includes the entire continuous space from the upper teeth to the postalveolar region.

schema in Figure 3.4 places the consonants and their vocalic equivalents in the same "space" of constriction location (from anterior/rostral to posterior/caudal) and frication intensity (none to substantial), and each class of vowel has a unique consonantal equivalent based on the articulators actually employed in creating the constriction. In Sūzhōu Chinese, the postalveolar vowels are transcribed as [i͡ʑ] and [y͡ʑ] for their [ʑ]-like quality and possible similarity to the series of consonant phonemes /ɕ tɕ tɕ^h dʑ/. Later, this dissertation will consider the question of whether a vowel and a consonant with similar but slightly different fricative noise production goals will tend to develop into a "series" of sounds that share motor programs in the ways discussed in the previous chapter.

3.3.4 Attestation of fricative vowels

The three major types of fricative vowels have different geographical distributions among the world's languages, which I briefly review here: starting with apico-alveolar, moving to postalveolar, and concluding with labial vowels (bilabial and labiodental). Apico-alveolar fricative vowels are broadly attested in a few cognate lexical sets across nearly all Chinese dialects, with the exception of most Min (e.g. Hokkien) and Yue (e.g. Cantonese) varieties (Lee and Zee, 2015). An unrounded apico-alveolar vowel [ɿ], for instance, is attested in 50 of the 70 Chinese varieties in a genetically balanced corpus collected by Zee and Lee (2007). Some non-Chinese languages in Qīnghǎi 青海 and Gānsù 甘肃 provinces of northwestern China have also developed apical vowels as allophones of /i/ when preceded by the expected fricatives and affricates, seemingly as an effect of contact with the local northwestern Mandarin dialects. Western Yugur, a Turkic language, is affected (Chen, 1986), as well as Amdo Tibetan (Wang, 2010) and several Mongolic languages: Minhe Mangghuer (Dwyer, 2008), Huzhu Mangghuer (Zhaonasitu, 1981), and Santa (Kim, 2003, p. 349; Field, 1997, p. 5).

Apico-alveolar fricative vowels are also attested beyond areas that speak, or are influenced by, Chinese. Numerous Tibeto-Burman languages spoken to the southwest of the historically Han-influenced area have been described with fricative vowels that appear to have developed independently of Chinese influence. These languages include the small Bai family (Dell, 1981), Yongning Na (Michaud, 2008, p. 176–77), Naxi (Michaud and He, 2015), many varieties of Yi (Li and Ma, 1983; Qumutiexi, 2010), and several Qiangic languages, including Lizu (Chirkova and Chen, 2013) and Ersu (Chirkova et al., 2015). Further afield, my fieldwork in Cameroon has also uncovered an instance of apical vowels in the Grassfields region of Cameroon. Oku, a language of the Ring group (Hyman, 1980), has a [ɿ]-like vowel that is an allophone of a mid central vowel /ə/ occurring after any alveolar syllable initial.

Most apico-alveolar fricative vowels have the "apical vowel phonotactics" described in the previous sections, occurring only after apico-alveolar fricatives and affricates such as /s, z, ts, dz/. In a handful of Chinese dialects, apico-alveolar vowels occur more freely, as in Héféi Chinese, where sequences such as [pɿ], [mɿ] are attested (Wu, 1995; Hou, 2009). If a language has apico-alveolar fricative vowels with "apical vowel" phonotactics and contrasts multiple places of articulation for coronal fricatives and affricates, it typically also has apico-postalveolar or postalveolar vowels with similarly restricted phonotactics. This situation is exemplified by Standard Chinese and other Chinese dialects that contrast apico-alveolar and "retroflex" apico-postalveolar fricatives and affricates, but is also commonly seen outside of Chinese proper, for instance in the Turkic, Tibetan, and Mongolic languages mentioned above.

Fricative vowels without the "apical vowel" phonotactics are most commonly postalveolar and are less frequently attested. These postalveolar fricative vowels are attested in two large discontinuous zones, mainly consisting of Chinese dialects: one covers northern Wú and southern (Jiānghuái 江淮) Mandarin in coastal Zhèjiāng and Jiāngsū provinces, and another covers northwestern Mandarin dialects in Qīnghǎi 青海 and Gānsù 甘肃 provinces (Wang, 2006; Zhao, 2007, p. 48; see also Zhu, 2015). Additionally, some varieties of Amdo Tibetan spoken in Qīnghǎi, adjacent to the aforementioned Northwestern Mandarin dialects, reportedly have a postalveolar fricative vowel (Wang, 2010), possibly as a contact phenomenon. Ersu, a Qiangic language, also has a postalveolar fricative vowel (Chirkova et al., 2015).

Unlike apico-alveolar vowels, postalveolar fricative vowels are also attested in a much wider variety of languages, including some well beyond China. They are attested in and around the Grassfields Bantu languages of northwestern Cameroon, specifically the Ring family within Grassfields Bantu (Faytak, 2017) and several adjacent non-Grassfields Bantu languages in contact with Ring (Fiore, 1987; Hyman, 1981; Connell, 2007). Babanki, another language of the Ring group, has rounded and unrounded postalveolar vowels with "apical vowel"-like phonotactics: they occur as allophones of /i/ and /ʉ/, respectively, when these vowels follow postalveolars /tʃ, dʒ, ʃ, ʒ/.

In Swedish, lamino-postalveolar vowels (possibly apico-postalveolar for some speakers) are attested as regional dialect variants of the usual /iː/ and occur with the relatively free phonotactics that would characterize that vowel. These are typically called the *Viby-i* or *Göteborges-i*, and are described as having a "buzzing", "dampened", or "dark" quality (Björsten and Engstrand, 1999; Engstrand et al., 2000), often simply called "Viby-coloring", a perceptual quality known to cued primarily by lowered F2 (Westerberg, 2016). There is growing evidence that Viby-colored variants are spreading into standard Swedish (Bruce, 2010, p. 136; Riad, 2013, p. 21; Westerberg, 2016, p. 102). The Viby-colored vowels in Stockholm and Göteborg Swedish have been the subject of some recent articulatory phonetic research suggesting that a variety of tongue shapes may correspond to the same "Viby-colored" perceptual variants. These tongue shapes appear to range from a bunched and convex shape that likely produces a postalveolar constriction (Schötz et al., 2011; Schötz et al., 2014; Westerberg, 2016) to a more complex double-bunched configuration (Westerberg, 2016).

Labial fricative vowels are somewhat less common than alveolar or postalveolar fricative vowels, but are broadly attested. Labial vowels realized with bilabial compression and with labiodental constriction are both attested; bilabially compressed vowels often exhibit bilabial trilling rather than frication. Both types of vowels tend to be phonemic and phonotactically unrestricted in the languages in which they occur. The Tibeto-Burman subfamilies of Naish, Loloish, and Bai have phonemic labiodental fricative vowels in addition to their aforementioned apical vowels (Dell, 1981; Michaud, 2008, p. 177; Michaud and He, 2015, p. 16), and Liangshan Yi is noted for its phonemic bilabial vowel (Li and Ma, 1983). Phonemic labiodental vowels are frequently attested in and around the Grassfields Bantu languages of northwestern Cameroon, including the languages Kom and Limbum (Faytak, 2017). Several dialects of the Fang language of northern Gabon are also recorded as having a labiodental fricative vowel as a variant of the more typical [u] (Medjo Mvé, 1997).

Both bilabial and labiodental vowels may also be present in the same language as allophonic variants of the same phoneme following bilabial stops and labiodental fricatives, respectively. This situation is common in Wú Chinese (Qian, 1992), including Sūzhōu Chinese, and also occurs in at least two Cameroonian Grassfields languages, Medʉmba (Olson and Meynadier, 2015) and Babanki (as determined through my own fieldwork); in each of these languages, the phoneme to which the labial fricative vowels belong is realized as a dorsal vowel without frication elsewhere ([ʉ] in the Grassfields languages; [əu] in Sūzhōu Chinese).

3.4 The postalveolar vowels in Sūzhōu Chinese

In this section, I review the small existing literature on Sūzhōu Chinese to develop a more detailed picture of the Sūzhōu Chinese “fricative vowels” (which, per the discussion in the previous section, are best referred to as postalveolar vowels). I begin by reviewing past work on the Sūzhōu Chinese postalveolar vowels, with some comparison to other vowels in Sūzhōu Chinese and other Wú Chinese varieties (Section 3.4.1). I then provide brief historical-comparative evidence for an origin of /i͡ʑ/ and /y͡ʑ/ from *i and *y, fairly canonical close dorso-palatal (i.e., high front) vowels, and speculate that the fricative noise targets inherent to /i͡ʑ/ and /y͡ʑ/ reflect the original functional motivation of the sound change (Section 3.4.2).

3.4.1 Prior research on the Sūzhōu Chinese fricative vowels

Sūzhōu Chinese is a typical northern Wú language in having several fricative vowels in its phonemic inventory: two apico-alveolar fricative vowels with the characteristic “apical vowel” phonotactics described in Section 3.3.1; a labial fricative vowel produced with bilabial compression after bilabial stops and labiodental frication after labiodental fricatives, and two postalveolar fricative vowels with less constrained phonotactics that are the subject of the experiment described later in this chapter. Both the apico-alveolar and the postalveolar fricative vowels come in pairs, each having a rounded and an unrounded version.

The range of articulations characteristic of the two postalveolar fricative vowels /i͡ʑ/, /y͡ʑ/ is described with some precision by Ling (2007) and Ling (2009). In particular, Ling (2009), from static palatography findings, infers that /i͡ʑ/ has either a dorso-postalveolar or lamino-postalveolar constriction, depending on the individual speaker; this is in contrast with /i/, which is dorso-palatal for all participants (pp. 42–45). These findings are consistent with a static palatography study I carried out in 2014 in the Berkeley PhonLab (Figure 3.5) with the goal of characterizing the equivalent vowels in several Wú dialects. The overall posture of the tongue in midsagittal cross-section can be inferred from Ling’s electromagnetic articulography studies of the same vowels for a handful of speakers (Ling, 2007; Ling, 2009, pp. 46–47): for all speakers examined, the highest point on the tongue appears to be more anterior than the dorsum, and the constriction location appears to be considerably more anterior than would be typical for for a cardinal [i].

Some articulatory and acoustic characterization of Sūzhōu Chinese’s apico-alveolar fricative vowels /ɿ/, [ʮ] can also be found in Ling (2009). Ling (2009)’s articulatory data on the apical vowels do not include static palatography, given that unlike fricative vowels apical vowels are obligatorily preceded by a fricative or affricate of the same constriction location—

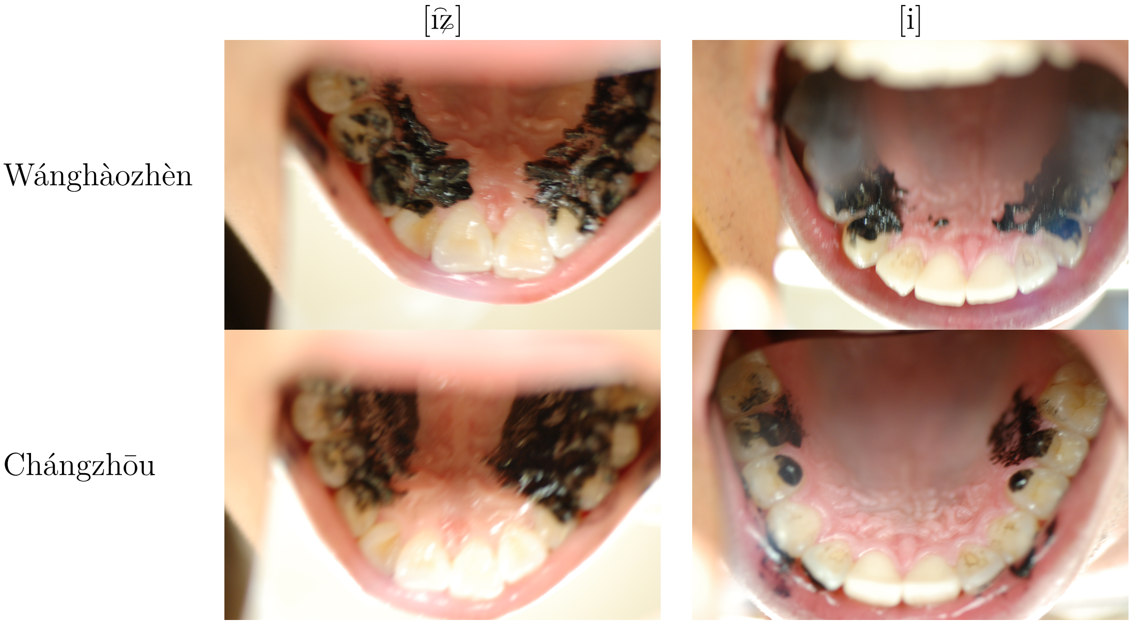


Figure 3.5: Unrounded postalveolar vowel [i͡ʑ] (left) and dorso-palatal vowel [i] (right) for speakers of (top to bottom) Wánghàozhèn 王浩镇 Chinese (Faytak, 2014) and Chángzhōu 常州 Chinese (own data), all northern Wú dialects in the same subfamily.

the contributions of the onset consonant and the apical vowel to overall tongue-palate contact cannot be disentangled. However, the EMA data for the apico-alveolar vowels (Ling, 2009, p. 52) are broadly in line with other articulatory descriptions of apical vowels in, e.g., Standard Chinese (Zhou and Wu, 1963; Lee-Kim, 2014; Faytak and Lin, 2015). The tongue is quite flattened and low, with the blade and tip of the tongue extended to touch the alveolar ridge. The articulation of the postalveolar vowels, the apico-alveolar vowels, and the dorsal-palatal vowels are all clearly distinct across subjects whose EMA data is examined.

Acoustically, Ling (2007) and Ling (2009) has found that productions of [i͡ʑ] for a given speaker of Sūzhōu Chinese generally have lower F2 compared to that speaker's [i]; the same relation applies to [y͡ʑ] and [y]. In the few other acoustic analyses of postalveolar vowels in the context of their broader vowel systems, on dialectal Swedish (Westerberg, 2016) and Len Mambila (Connell, 2007), all describe a similar pattern for F2: lower for [i͡ʑ] than for [i].[6] Sūzhōu Chinese's apical vowels have an even lower F2, by about 900 Hz across subjects compared to the postalveolar vowel matching for roundness. Ling offers an explanation for the depressed F2 values typical of /i͡ʑ, y͡ʑ/ and the apical vowels: in an idealized resonator tube model of vowel production (Fant, 1960, pp. 76–77; Stevens and Keyser, 1989), F2 is predicted to increase as a lingual constriction advances forward along the palate, but after

---

[6] In Westerberg (2016), the Swedish vowel that acoustically resembles cardinal [i] is typically understood as /eː/, which appears to have raised for her subjects. Both /eː/ and the [i͡ʑ]-like vowel that is the object of comparison exhibit a comparable F1 typical of high vowels.

a certain point, as the constriction continues to advance and the back tube continues to lengthen, F2 decreases as the affiliation of F2 switches from the front resonating cavity to the back resonating cavity (2009, pp. 49–57).

In some cases, Ling's participants also exhibit a higher F1 in their postalveolar vowels relative to their dorso-palatal vowels; the apical vowels generally exhibit an even higher F1 compared to both other sets of vowels. Ling's explanation for this pattern is more complex: in an [i]-like tube configuration, F1 is typically derived from the Helmholtz resonance of the back cavity, since the back cavity is effectively closed by the constriction between the tongue and palate. The frequency of the Helmholtz resonance f is inversely related to the length of this constriction l_c:

$$f = \frac{c}{2\pi}\sqrt{\frac{A_c}{A_b l_b l_c}}$$

When l_c is relatively short, as it is for apical and laminal vowels, F1 is relatively high (Ling, 2009, pp. 55–57). Ladefoged and Lindau (1989) find that moving a modeled constriction location forward on the tongue results in a gradual increase in F1, as well as the expected gradual decrease in F2. I speculate that smaller values of l_c occur in more anterior vowels due to the fact that the curve of the palate brings the alveolar/postalveolar region more perpendicular to the anterior portions of the tongue, which can make a relatively short constriction against the palate due to this angle of attack and the greater mobility of the tongue blade and tip compared to the dorsum.

Ling (2007, 2009:22-23) also measures the harmonic-to-noise ratio (HNR) of the dorso-palatal, postalveolar, and apico-postalveolar vowels using Praat's HNR algorithm (Boersma and Weenink, 2017). A harmonic-to-noise ratio of a non-pathological, modally voiced vowel is typically above 20 across the spectrum. HNR for [i͡ʑ, y͡ʑ] is lower than 10 dB in the spectral band above 2 kHz, indicating a substantial amount of aperiodic energy for these vowels that is consistent with impressionistic assessments (Qian, 1992; Hu, 2007). HNR is also below 10 dB above 2 kHz for [ɿ, ʮ], indicating a substantial amount of aperiodic energy in this spectral range. By way of comparison, the HNR of a voiced strident fricative is typically in the range of 5–10 dB, and a voiced non-strident fricative is typically in the range of 10–15 dB (Maniwa et al., 2009, p. 3971).

3.4.2 Historical-comparative evidence for dorsal origins

The postalveolar and alveolar fricative vowels in Sūzhōu Chinese have developed over very different scales of time and space. Apical vowels are present in cognate lexemes in most dialect groups of Chinese, suggesting that an initial "apicalization" change can likely be traced back to Middle Chinese (Chen, 1976; Yu, 1999), thought to be the common ancestor of all varieties of Chinese other than Min. In Wú Chinese, the apico-alveolar vowels [ɿ, ʮ] are present in all dialects described in Qian (1992), and a [ɿ]-like apical vowel category *ɨ

PWu	*-i	*-jen	*-y	*-ɥɤn
Sūzhōu 苏州	-i$_z$	-i	-y$_z$	-y ∼ -yø
Dānyáng 丹阳	-i$_z$	-ɪ	-y$_z$	-ʏ
Chángzhōu 常州	-i$_j$	-ɪ̃ ∼ -ɪ	-y$_y$	-iɔ
Chóngrén 崇仁	-i$_z$	-iẽ	-y$_y$	-yœ̃
Jiāxìng 嘉兴	-i	-ie	-y	-yɤə
Tàipíng 太平	-i	-iẽ	-y	-yœ̃

Table 3.3: Modern Wú reflexes of selected Proto-Wu (PWu) rhymes. Modern Wú transcriptions from Qian (1992); PWu sets from Ballard (1969).

is in fact reconstructed for Proto-Wú (Ballard, 1969, p. 68). Some instances of [ʅ], and apparently all instances of [ɥ], in the modern dialects are reflexes of vowels in *si/*ɕi or *sy/*ɕy sequences (respectively) which have developed earlier, and in a much wider range of languages, than the fricative vowels [i͡ʑ, y͡ʑ] (Ballard, 1969).

The postalveolar fricative vowels, on the other hand, have developed only in scattered pockets around China, with one such development affecting Wú and nearby Jiānghuái Mandarin relatively recently. Wú and Jiānghuái Mandarin do not form a proper clade within Chinese, but are immediate geographic neighbors across the Yangtze River, suggesting that this innovation spread from one family to the other. However, the directionality of influence in this case is not clear. A second, less-studied large area consisting of Northwestern Mandarin and some neighboring non-Chinese languages also exhibits [i͡ʑ] as a reflex of *i (Wang, 2006; Wang, 2010). This area and other, smaller areas are discussed further in Zhu (2004) and Zhao (2007).

In Wú, the chronology of the change of *-i and *-y to modern reflexes [i͡ʑ] and [y͡ʑ] can be inferred relatively well from existing comparative data and from Ballard (1969)'s reconstruction of Proto-Wu (PWu). As PWu *-i and *-y undergo a constriction location change to [i͡ʑ] and [y͡ʑ] in most Wú dialects, a pair of PWu rhymes *-jen and *-ɥʌn raise, denasalize, and monophthongize in a variety of dialect-specific developments. These two changes have been argued to constitute a sort of push chain shift, with the raising and monophthongization of *-jen and *-ɥʌn driving the fricativization and place change of *-i and *-y (Zhu, 2004, p. 448). However, Wú comparative data from Qian (1992) (Table 3.3) show that fricativization occurs largely independently of raising and monophthongization of the lower rhymes, in some cases without any acoustic proximity between the higher and lower rhymes that could drive a push chain (in i.e. Chóngrén 崇仁, Chángzhōu 常州). This suggests that raising of *-i and *-y preceded developments in *-jen and *-ɥʌn, making this a short *pull* chain shift. Under this version of events, fricativization of *-i and *-y leads, and the lower, diphthongal, nasal reflexes of *-jen and *-ɥʌn sporadically raise, monophthongize, and denasalize, coming to occupy the acoustic space formerly occupied by *-i and *-y.

This regular pattern (in northern Wú and other areas) can be explained by listener-driven sound change due to misperception. The initial development of contrastive frication in reflexes of *-i and *-y may be triggered by phonologization of noise that is incidentally produced in high vowels (Faytak, 2014). One could argue that this is a listener-driven sound change, primarily due to hypocorrection (Ohala, 1993). High (dorsal) vowels occasionally exhibit some fricative noise; a fricative noise source can easily be generated given the high rate of airflow required to initiate voicing and the narrow vocal tract aperture (Shadle, 1990). Partially devoiced or fricated high vowels are amply attested as phonetic variants of high vowels (Maekawa and Kikuchi, 2005; Fagyal and Moisset, 1999; Smith, 2003). Listeners may incorrectly identify the frication as a phonetic target rather than an incidentally produced attribute of that vowel.

Regardless of whether or not a push chain shift can be said to have occurred, Sūzhōu Chinese's especially crowded vowel space may play a role in the development of frication and its subsequent changes. If numerous vowels cluster in the same region of the formant frequency space, as appears to be the case for Wú in general and Sūzhōu Chinese in particular, then formant frequency loses its reliability as a cue for accurately distinguishing these vowel categories. In light of this reduced cue reliability, speakers of Sūzhōu Chinese might be expected to engage in "probabilistic enhancement" (Kirby, 2013) of fricative noise, a more reliable cue to the higher and more constricted *-i, *-y vowels when compared to the lower, possibly nasalized *-jen and *-ɥʌn rhymes.

3.4.3 Conclusion

The previous sections confirm that postalveolar fricative vowels as reflexes of Proto-Wú *-i and *-y are well-established in Sūzhōu Chinese, with most speakers exhibiting at least some frication and a different perceived vowel quality from their dorso-palatal /i/ and /y/. As discussed previously, perceptual bias toward using informative cues to category has likely driven the phonologization and enhancement of fricative noise as a cue to these categories. Of note for the rest of this dissertation, however, is that this process does not determine what specific articulatory strategy a speaker employs for producing this fricative noise. What interests us for the purposes of the experiment described in the following chapter is the specific articulatory strategies speakers have come to employ for fricative noise generation and how they relate to other articulatory strategies in the inventory, namely those typical of fricative/affricate series like the palatoalveolars.

The situation described above hints at the primacy of "speaker-side" factors in changing the articulatory strategies used for the Sūzhōu Chinese fricative vowels: however, most other sources of variability and change first have to be ruled out. On one hand, contact effects can be ruled out. The change from *-i, *-y to fricative vowels and the ongoing move within the fricative vowel categories to a relatively anterior active articulator do not appear to be due to contact with Standard Chinese, in spite of the multilingual situation described in Section 3.1. Standard Chinese does not exhibit postalveolar vowels in the phonotactic environment in which they have developed in Sūzhōu Chinese. While Standard Chinese and Sūzhōu Chinese do both exhibit apico-alveolar fricative vowels following apico-alveolar fricative and affricate consonants, these segments are known to be a shared inheritance (see Section 3.4.2).

Contact effects owing to contact with Standard Chinese actually appear to erode the distinctiveness of the postalveolar fricative vowels. Sūzhōu speakers with stronger Standard Chinese proficiency sometimes the contrasts between the pairs /i/–/i͡ʑ/ and /y/–/y͡ʑ/ altogether, merging to the dorso-palatal vowel in each pair ([i], [y]), which happen to be the vowels in the Mandarin cognate words. This same innovation has become the norm among the youngest generations of Shànghǎi dialect speakers (Zhu, 2006), affecting the same cognates to Standard Chinese and having the same outcome.

Social factors within the Sūzhōu Chinese community can also more tentatively be ruled out as driving this relatively subtle change, making listener-driven explanations less likely to have substantial explanatory power. In general, fricative vowel realization does not appear to be a sociolinguistically meaningful variable. Although there may well be measurable differences between, for instance, the [i]-like conservative formant values and the [ɨ]-like innovative formant values for /i͡ʑ/, neither variant appears to rise to the level of conscious awareness for most Sūzhōu Chinese speakers, and neither variant is associated with a coherent stereotype. This observation, of course, is not based on systematic investigation. Further study is needed to determine whether or not Sūzhōu Chinese speakers can, in fact, distinguish between the various realizations of /i͡ʑ/ and assign social meaning to their use.

Chapter 4

Articulatory reuse in the Sūzhōu Chinese fricative vowels

This chapter presents an ultrasound tongue imaging experiment that evaluates the uniformity of articulatory strategies used to produce a set of speech sounds in Sūzhōu Chinese. Recall that in Chapter 2, a theory of articulatory reuse was developed. During L1 acquisition, when learners have not yet developed robust motor-articulatory-acoustic mappings (McAllister Byun et al., 2016), stored controls used to achieve acoustic goals with some consistency may generalize to newly created motor programs, possibly through trial-and-error learning (Loeb, 2012). I theorize that this reuse due to behavior exhibited during skill learning generally has some degree of impact on adult speech production strategies in the form of articulatory uniformity. Uniformity of articulatory strategies across a series of speech sounds sharing a similar phonetic goal is amply attested (Flege, 1982; Keating, 2003; Ménard et al., 2008; Chodroff and Wilson, 2017; Chodroff, 2017); this constrained variation in the phonetic realization of phonological primitives can be attributed to a "yoking together" of the articulatory means used for speech-directed tasks as a result of reuse during learning.

Here, the set of speech sounds at issue are the two postalveolar fricative vowels of Sūzhōu 苏州 Chinese, transcribed in Chapter 3's detailed description as /i͡ʑ/ and /y͡ʑ/. These fricative vowels provide a unique opportunity to investigate patterns of articulatory uniformity that are theorized to result from articulatory reuse, given that their fricative noise targets, which spectrally resemble [ɕ] and and other alveolopalatal fricatives and affricates, may be produced either using a [ɕ]-like articulatory strategy or via some other lingual configuration without causing a perceptible, consistent difference in acoustic output. The experiment in this chapter therefore examines the lingual articulation of the relevant fricative vowels and consonants for a large and representative sample of Sūzhōu Chinese speakers ($n = 44$), in an effort to ascertain the extent to which the fricative vowels /i͡ʑ/ and /y͡ʑ/ are produced similarly to fricative consonants such as /ɕ/.

The findings reported here suggest that most individual speakers of Sūzhōu Chinese habitually favor an articulatory strategy for the fricative vowels /i͡ʑ/ and /y͡ʑ/ that is uniform with their articulatory strategy for /ɕ/. That is, both the fricative consonants and the fricative vowels have an articulatory strategy that uses essentially the same lingual posture to produce a lamino-postalveolar constriction that, like /ɕ/, could be described as alveolopalatal. However, some speakers' preferred production strategies for /i͡ʑ/ and /y͡ʑ/ are not uniform with /ɕ/ or other vowels such as /i/ and /y/. The choice of uniform or non-uniform production strategies appears to be largely idiosyncratic, and speakers' responses to coarticulatory pressures from adjacent fricative consonants also vary substantially and idiosyncratically.

I arrive at this understanding of the speaker population through two complementary analyses of ultrasound video data. First, I demonstrate with a data-driven qualitative analysis of reduced-dimensionality representations of ultrasound video data that a pattern of uniformity prevails across the study population. Second, a quantitative assessment of a metric of relative /ɕ/- and /i/-likeness, derived from ultrasound video data via dimensionality reduction, shows that variables such as age and gender do not predict individuals' articulatory strategies for /i͡ʑ/ and /y͡ʑ/, although presence of a /ɕ/ in the immediate segmental context has a small coarticulatory effect on similarity to /ɕ/. Nonetheless, the effect of coarticulation is small, suggesting a relatively invariant tendency to reuse /ɕ/-like tongue configurations even in the absence of coarticulatory influence from /ɕ/, modulo speaker idiosyncrasies.

These findings are in keeping with existing case studies of uniformity and "structured variation" (e.g., Keating, 2003; Ménard et al., 2008; Chodroff, 2017), which generally find a strong tendency toward uniformity of outputs in some phonetic sense (articulatory, acoustic, or both) where one need not necessarily exist; this tendency toward uniformity along the phonetic dimension(s) examined is modulated by individual differences in the degree of uniformity. However, the Sūzhōu Chinese case examined here is also a case of covert articulatory uniformity—the difference between uniform and non-uniform articulatory strategies is not readily perceptible—pointing to the need for further investigation of uniformity through direct articulatory assessment rather than indirect acoustic indices of articulation.

The structure of this chapter proceeds as follows: Section 4.1 provides a brief overview of the hypotheses evaluated in this experiment and pilot data that informed the generation of these hypotheses. Sections 4.2–4.3 detail the study population, experimental methods, the means of acoustic analysis, and the means of dimensionality reduction and analysis of the ultrasound image data. Section 4.4 is given over to a preliminary exploration of the acoustics (Section 4.4.1), overall characterization of the lingual posture of the apico-alveolar and postalveolar fricative vowels (Section 4.4.2), and a more specific examination of the similarities of the postalveolar vowels /i͡ʑ/ and /y͡ʑ/ to /ɕ/ in terms of lingual posture (Section 4.4.3). Discussion of results and concluding notes follow in Section 4.5.

4.1 Experiment background and aims

Before discussing the hypotheses considered by the main ultrasound experiment, which begins in Section 4.2, I first provide an overview of exploratory pilot data that has contributed to the formation of the specific hypotheses investigated. Following this primer in Section 4.1.1, the hypotheses themselves are laid out in Section 4.1.2.

	Alveolar "series"	Postalveolar "series"
Affricate	ts tsʰ	tɕ tɕʰ dʑ
Fricative	s z	ɕ
Fricative vowel (unrounded)	ɿ	i͡ʑ
Fricative vowel (rounded)	ʮ	y͡ʑ

Table 4.1: Sūzhōu Chinese sounds involving the production of strident frication. All segments except the rounded apico-alveolar fricative vowel [ʮ] exhibit phonemic contrast on distributional grounds; [ʮ] is analyzed as an allophone of /y͡ʑ/ that occurs following alveolar fricative/affricate consonant onsets.

4.1.1 Sūzhōu Chinese overview and pilot ultrasound data

The phonological inventory of Sūzhōu Chinese is rich in segments whose production involves fricative noise generated at a coronal place of articulation. These include two series of coronal fricative/affricate consonants, one alveolar or dental and the other alveolopalatal, and four so-called FRICATIVE VOWELS that have been described as alveolar or postalveolar in constriction location (Wang, 1987; Ye, 1988; Qian, 1992; Ling, 2009). This section of the inventory is shown in Table 4.1, which summarizes material presented in Chapter 3. Speakers appear to produce the fricative vowels with sibilant or shibilant fricative noise (depending on anteriority of constriction) that is not obviously laryngeal in origin; that is, all are modally voiced (Ling, 2009). This fricative noise also does not appear to be attributable to coarticulation with other segments or a high rate of speech, as shown in Figure 4.1: an onsetless syllable uttered at a normal rate consisting of /i͡ʑ/ differs from one consisting of /i/ in several respects, including the presence of substantial aperiodic energy above 4 kHz that is absent in /i/.

The above data are suggestive of two series of sounds sharing similar lingual postures and encompassing both consonants and vowels; each series possibly also shares the production of a particular type of fricative noise as a target. To explore this tentative hypothesis further, I collected pilot data on the fricative vowels and consonants of Sūzhōu Chinese. Participants were recruited from the UC Berkeley campus population, and recording took place in a sound-attenuated booth in the UC Berkeley PhonLab. Participants were compensated $10 for their time and effort; all procedures described here were approved by the UC Berkeley IRB. Tongue ultrasound imaging in midsagittal section was recorded at 107 fps on an Ultrasonix SonixTablet using a C9-5/10 microconvex transducer held in place by an Articulate Instruments Ltd. stabilization headset (Articulate Instruments Ltd., 2008). Audio synchronized with the ultrasound video was recorded with an AKG 535 EB microphone and digitized through a Steinberg UR22 USB audio interface (see Section 4.3.3 for synchronization method).

Frames from the acoustic midpoints of relevant segments (/i͡ʑ/, /i/, and /ɕ/) were se-

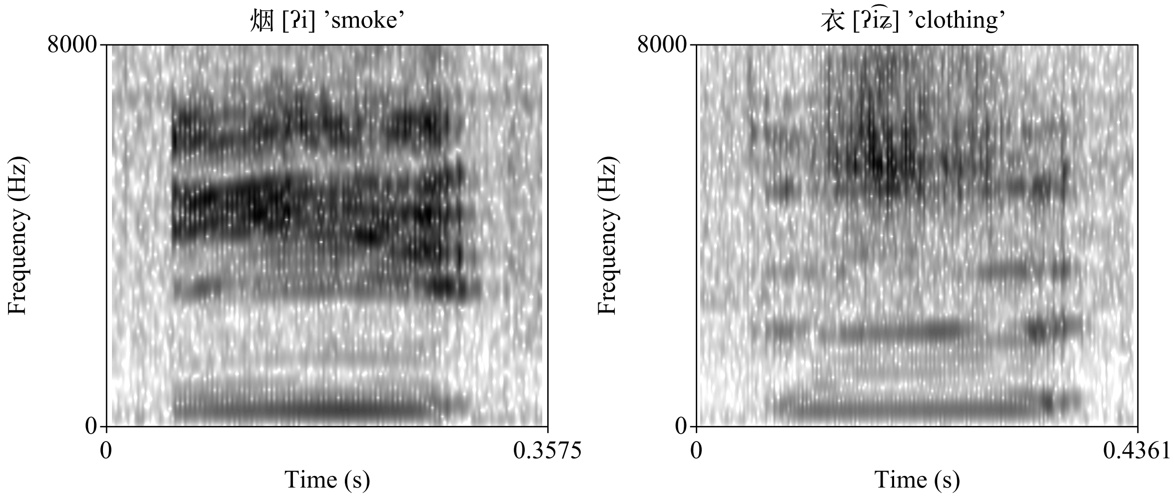


Figure 4.1: Spectrograms of utterances of [ʔi^{44}] ‘smoke’ (left) and [ʔi͡ʑ^{44}] ‘clothing’ (right) from a female speaker of Sūzhōu Chinese (Speaker 3). Note fricative noise above 4 kHz in the [i͡ʑ] in ‘clothing’ and reduced amplitude of the upper formants, as well as the large difference in F2 relative to the [i] in ‘smoke’.

lected using acoustic landmarks in the time-aligned audio, and tongue surface contours were extracted using EdgeTrak (Li et al., 2005). Each speaker’s collected contours were submitted to a smoothing-spline ANOVA (SSANOVA) model to estimate the speaker’s typical tongue position for each segment (Gu, 2002; Davidson, 2006); a polar coordinate system was used, after Mielke (2015).

The results of the pilot study are largely in keeping with Ling (2009)’s description based on electromagnetic articulography and static palatography, which finds evidence consistent with two covertly different types of constriction for speakers’ /i͡ʑ/: dorso-postalveolar and lamino-postalveolar (pp. 42–45). The present data offers some advantages over Ling (2009), namely the ability to image the entire tongue surface for a larger number of participants. In Figure 4.2, Speakers 01, 05, and 07 are seen to each use various articulations that, in terms of the activity of the independent lingual articulators, are similar to their particular productions of /ɕ/; the fricative and the vowel could both be described as lamino-postalveolar. The degree of similarity is particularly striking for Speakers 01 and 05. Speaker 08, on the other hand, exhibits a non-uniform dorso-postalveolar variant, somewhat akin to a hyperarticulated /i/: the entire tongue dorsum bulges up and forward in the direction of the postalveolar area.

Taken together, all of this suggests a tendency toward articulatory reuse of a /ɕ/-like articulatory strategy in all sounds with [ɕ]-like fricative noise. Speaker 08 presents an exception to this tendency in having an articulatory strategy for producing /i͡ʑ/ that is uniform neither to the fricative consonant /ɕ/ nor the vowel /i/. It is also of note that Speaker 08 is much older than the other speakers, suggesting that uniformity is an innovative feature and may be age-graded. The ultimate goal of the remainder of this chapter is thus to evaluate

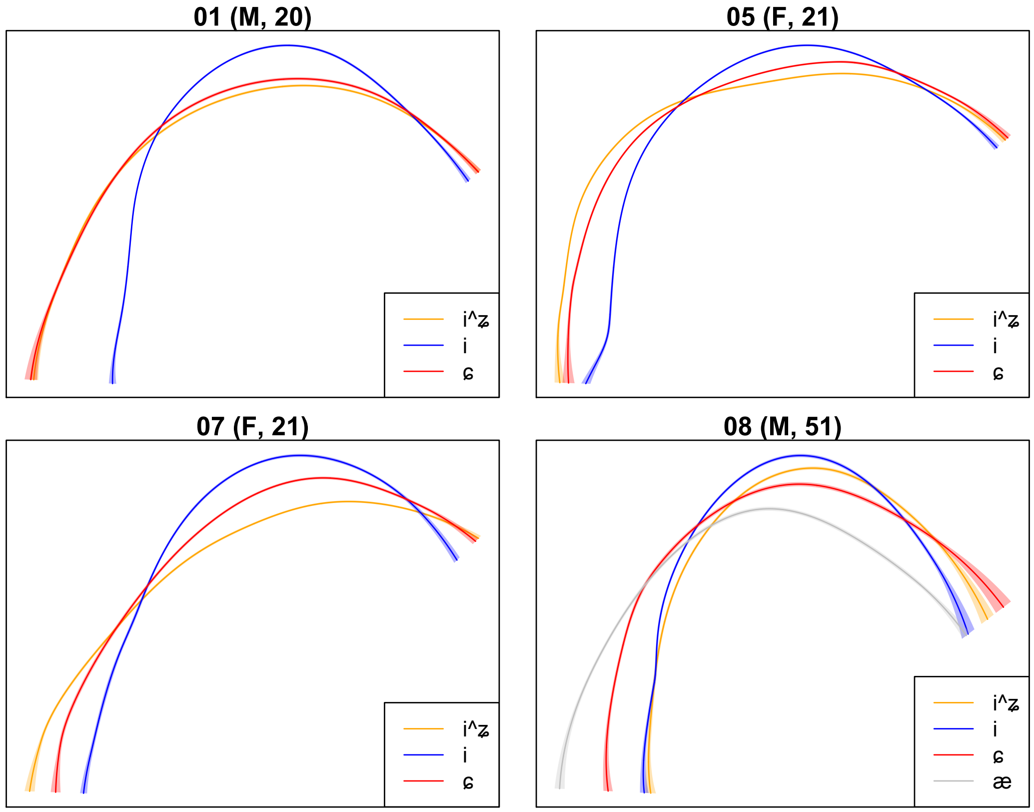


Figure 4.2: Tongue position for the postalveolar vowel /i͡ʑ̥/ (i^ʑ̥ in legends), the dorso-palatal vowel /i/, and the fricative /ɕ/ in four Sūzhōu Chinese speakers. Smoothing-spline ANOVA estimates of tongue contour position with 95% confidence intervals from ultrasound images are shown. Right is anterior.

the extent to which speakers' articulatory strategies are uniform among their fricative vowels and their fricative consonants, and whether this tendency changes with respect to variables such as age.

4.1.2 Hypotheses

The experiment on the Sūzhōu Chinese fricative vowels in this chapter tests a variety of hypotheses related to uniformity of motor programs across segments with a set of closely related acoustic goals. Two groups of segments with similar acoustic goals are investigated: the apico-alveolar fricative /s/ and the apparently apico-alveolar vowels /ɿ/, [ʮ][1] on one hand, and the lamino-postalveolar fricative /ɕ/ and the apparently dorso- or lamino-postalveolar vowels /i͡ʑ̥/, /y͡ʑ̥/ on the other hand, focusing particularly on the lamino-postalveolar group.

---

[1] Recall that [ʮ] is analyzed here as an allophone of /y͡ʑ̥/ on distributional grounds.

I look for uniformity at two distinct levels: as a characteristic of the individual speaker and as an attribute of the broader population.

Hypothesis 1a relates to uniformity within individual speakers' motor program repertoires:

> Hypothesis 1a: a speaker will articulate the vowels described as "lamino-postalveolar" (/i͡ʑ/ and /y͡ʑ/) with tongue posture targets indistinguishable from the tongue posture for the lamino-postalveolar fricative /ɕ/.

Based on inter-speaker variability observed in prior research and my own study of Sūzhōu Chinese, it is not expected that Hypothesis 1a will be confirmed for all speakers. As discussed in Chapter 2, speakers may respond differently to uniformity (and other) pressures in language development and adult maintenance, resulting in different uniformity outcomes in their individual speech motor program repertoires. This is reminiscent of the different speaker groups noted in Keating (2003), who idiosyncratically prioritize uniformity of acoustic outputs or articulatory strategies in their production of English short-lag VOT stops. Accordingly, on one hand, I expect that only some Sūzhōu Chinese speakers will prioritize uniform motor programming, producing their fricative vowels and (voiceless) fricative consonants with essentially identical tongue positions. This is the case for some participants in the pilot experiment as shown in Figure 4.2 (Speaker 01, and to some extent Speaker 05).

As discussed in Chapter 2, speakers may also, on the other hand, prioritize other factors in organizing their speech production, such as aerodynamic optimization: articulating voiced fricated sounds efficiently and consistently may require articulations distinct from their voiceless fricative consonant counterparts. Maintaining both voicing and frication simultaneously is difficult due to the double air pressure drop required along the vocal tract: across the glottis, and then across the supralaryngeal constriction (Catford, 1977; Ohala, 1983). Seemingly as a result of this constraint, voiced and voiceless segments are typically implemented with subtle differences, including passive pharyngeal expansion in voiced obstruents (Proctor et al., 2010) and (more relevant for the present case) small differences in tongue posture or constriction degree (Faytak and Lin, 2015).

This would presumably lead some Sūzhōu Chinese speakers to have slightly different tongue postures for the consonant-like and vowel-like sounds within each place of articulation. These adjustments are expected to be slight changes to constriction degree or cavity volume, rather than qualitatively distinct tongue shapes. The pilot data discussed above suggests that where differences do exist between /ɕ/ and the fricative vowels, this is typically the case: while Speakers 05 and 07 exhibit a range of strategies for /i͡ʑ/ in which the tongue root is retracted further relative to /i/ than the consonant /ɕ/ and the dorsum lowered relative to /ɕ/, the difference is one of degree rather than type. The slight enlargement of the cavity behind the fricative constriction likely facilitates the production of voiced frication. Speaker 08, who is more linguistically conservative[2] than the other three speakers, instead has an

---

[2] For instance, Speaker 08 does not palatalize /s/ before the vowels /i/ and /i͡ʑ/, a hallmark of a more innovative Sūzhōu Chinese phonology; the three younger pilot participants all palatalize.

/i͡ʑ/ posture which is similar to /ɕ/ in its anterior portion (having an anteriorly protruded tongue blade) but similar to /i/ elsewhere along the contour (having a very convex overall shape and an advanced tongue root). One could describe this strategy as a sort of contrastive hyperarticulation of /i/, in contrast to the /ɕ/-like tongue shapes used by the other speakers.

Restrictive phonotactics commonly apply to fricative vowels: for instance, apico-alveolar [ɿ] in Standard Chinese must be preceded by a homorganic onset consonant from the set /s, ts, ts^{h}/ (Lee-Kim, 2014; Faytak and Lin, 2015), making it difficult to disentangle possible articulatory uniformity of /s/ and [ɿ] from coarticulatory influence of /s/ on [ɿ]. While these phonotactics also apply to the Sūzhōu Chinese apico-alveolar fricative vowels /ɿ/ and [ʮ], the dorso-postalveolar or lamino-postalveolar vowels in Sūzhōu Chinese are relatively unrestricted, occurring both with homorganic fricative onsets (such as /ɕ/) and without them (/p/, /f/, Ø, etc.). The extent to which a speaker produces /i͡ʑ/ and /y͡ʑ/ similarly to /ɕ/ across /ɕ/-initial and /ɕ/-free contexts is an important gauge of the source of the uniformity that may be observed, allowing coarticulation to be taken into account or to be ruled out as a contributor to uniformity altogether. Thus, in addition to Hypothesis 1a, we also consider Hypothesis 1b at the individual level:

> HYPOTHESIS 1B: There will be no significant difference between the similarity of the fricative vowels /i͡ʑ/, /y͡ʑ/ to /ɕ/ in two contexts: contexts where /ɕ/ is present and exerting coarticulatory pressure, i.e. /ɕi͡ʑ/, and contexts where /ɕ/ is not present, i.e. /pi͡ʑ/.

It also stands to reason that, with successive generations of transmission, the biases introduced by exploratory, trial-and-error learning (as discussed in Chapter 2) that lead to reuse and (eventually) uniformity in adult inventories alos ought to result in a narrowing of the attested articulatory strategies for the developing lamino-postalveolar vowel. Articulatory similarity of /i͡ʑ/ and /y͡ʑ/ to /ɕ/ at the population level may thus exhibit a relationship with respect to speaker age. Successive generations of adult speakers of a language may show an increasingly close similarity between the already-similar fricative vowels and fricative consonants as apparent time passes:

> HYPOTHESIS 2: The speaker population will have an increasingly /ɕ/-like articulatory target for the lamino-postalveolar vowels /i͡ʑ/, /y͡ʑ/ in apparent time.

The fact that the younger speakers in the pilot data appear to have converged around a /ɕ/-like (or hyper-/ɕ/-like) strategy may suggest a winnowing of attested articulatory strategies in the younger generation. This may be in contrast to Speaker 08 and other speakers from older generations, which, having had fewer generations of "exposure" to uniformity in the linguistic environment, may collectively use a wider range of articulatory strategies for /i͡ʑ/.

Exploration of these hypotheses will primarily proceed using articulatory data extracted from ultrasound video. Acoustic characterization of the Sūzhōu Chinese vowels will be largely exploratory, especially given that the specific nature of the articulation of /i͡ʑ/ and /y͡ʑ/ for a given speaker is difficult to recover from acoustic data alone, due in part to the atypical articulatory-acoustic relations described for these sounds by Ling (2009).

4.2 Experimental method

All recruitment and experimental procedures described in this section and the following sections were approved by the UC Berkeley IRB. All participants provided informed consent, signed a media records release, and volunteered demographic information using survey forms translated into Standard Chinese; consent forms and survey instruments are provided in Appendix A.

4.2.1 Participants

Participants were recruited from the population of native speakers of Sūzhōu Chinese living in and around the city of Sūzhōu. A total of 44 participants were recruited (28 F, 16 M, ages 18–57, mean age 34.7). Subjects are largely homogenous in residential and linguistic history. All are long-term residents of one or more of the urban districts of Sūzhōu city, and nearly all participants report native-like proficiency in both Sūzhōu Chinese and Standard Chinese. All speak Sūzhōu Chinese at a native-like level, and all also speak Standard Chinese with varying levels of self-rated ability; all participants reported having learned some amount of Mandarin between the ages of 5 and 7, in their first years of primary school. A minority, mainly speakers younger than the age of 30, reported some ability in English; few participants reported proficiency in any other language. Selected metadata for participants, including residential history, linguistic background, age, and gender, can be found in Appendix B.

Participants were recruited through snowball sampling (Goodman, 1961) from initial contacts. During debriefing after completing the study, participants were asked to refer between three and five other potential participants of any age or gender who were interested in the research and sufficiently fluent speakers of Sūzhōu Chinese. The majority of participants did not generate new referrals, and several who did generate referrals generated more than five. Participants were paid ¥100 (about $15 by the USD-RMB exchange rate at the time of the experiment) as compensation for their time and effort; with study procedures lasting about 1.5 hours, this works out to an average hourly rate of $10 per hour.

4.2.2 Location and recording apparatus

Recordings consisted of three types of records: ultrasound tongue imaging video, audio, and video of the participant's face. Face video records are not used in the data set to be discussed in this chapter but are nonetheless detailed below. Recording was carried out in two locations in April of 2017. Four participants (Speakers 1, 2, 3, and 44) were recorded in a sound-attenuated booth in the Department of Chinese Language and Linguistics at Fudan

University in Shànghǎi. The remaining 40 were recorded in a quiet hotel room rented for this purpose in Gūsū district, Sūzhōu city.

Ultrasound video was recorded using an Echo Blaster portable ultrasound device equipped with a PV6.5/10/128 Z-3 microconvex probe. The frame rate for these recordings was typically 54 fps. The ultrasound probe was held in place under the chin using an Articulate Instruments, Ltd. stabilization headset (Articulate Instruments Ltd., 2008); the headset was adjusted to maximize freedom of movement of the mandible while maintaining full contact of the probe with the participant's skin. Given the need to accommodate differences in participant jaw and chin morphology, the probe angle relative to the occlusal plane varies from participant to participant.

Audio recordings were made with a Sony ECM-77B electret condenser microphone clipped to the arm to which the helmet's right cheekpad is attached. This typically resulted in a microphone position 1–2 inches to the right of the midline laterally and approximately an inch above the mouth. Audio recordings were made at a sampling rate of 44.1 kHz and digitized using a Focusrite Scarlett 2i2 USB audio interface. The USB audio interface was configured to accept the recorded speech signal and the pulse train automatically generated by the Echo Blaster. These channels were joined into a single stereo recording, to enable synchronization of acoustic landmarks with the recording of particular ultrasound frames to a high degree of precision.

Video data were also collected, but are not used further in the analyses discussed in this thesis. Video recordings were made with a tripod-mounted Zoom Q4 Handy video recorder. The video recorder's microphone was not used to record audio; rather, the ultrasound device's pulse train was split and fed into the video camera's external audio input, resulting in a two-channel recording in which the pulse train signals that the ultrasound device is recording frames. Inclusion of the pulse train as an added channel to both the audio and video data allows for easy alignment of both channels to the ultrasound signal and to each other in future research on Sūzhōu Chinese that may make use of this data set.

The experimental setting for most of the participants, a rented hotel room, introduced occasional background noise[3] and slight echo, but recordings taken in this setting are otherwise unremarkable and suitable for acoustic analysis. Video recordings, although not discussed further here, can be used to exemplify the setting and the apparatus as positioned on a participant's head; two frames from one such video are given in Figure 4.3.

4.2.3 Stimuli

Subjects were instructed to produce 26 simplified Chinese characters (hànzì) displayed on a laptop screen with a Sūzhōu Chinese reading (as opposed to a Standard Chinese reading). The stimuli and their expected Sūzhōu Chinese readings are shown in Table 4.2. Stimuli include a series of fricative-initial stimuli, with onset /s/ or /ɕ/, a set of stimuli without

---

[3] While noise was mostly sporadic (door slams, a clarinet being practiced) there were also some sustained, aperiodic background noises that could negatively affect the precision of holistic spectral measurements like harmonic-to-noise ratio (the room's climate control system, vacuum cleaners).

Figure 4.3: Two views of the ultrasound stabilization helmet, ultrasound probe, and microphone (clipped to right cheekpad, most easily observed in image at right) as fit to one participant.

onsets, and a series of stimuli with onset /p/. These onsets are combined with the lamino- or dorso-postalveolar vowels /i͡ʑ/ and y͡ʑ/, the apico-alveolar vowels /ɿ/ and [ʮ], and the bilabial and labiodental vowels [ə͡ß] and [ə͡v][4] as Sūzhōu Chinese phonotactics permit, with the addition of the high front (dorso-palatal) vowels /i/ and /y/ for comparison. Each of the 26 stimulus characters was read ten times, for a baseline total of 260 productions per participant. The actual total number of trials recorded by each participant varies, owing to several related factors described below. In particular, variant readings for stimuli occasionally did not contain any segments destined for analysis, and timing errors and disfluencies occasionally eliminated usable stimuli from productions altogether.

A major factor leading to variation in the number of productions recorded is the mild difficulty some participants had with assigning consistent Sūzhōu Chinese readings to the stimuli. Sūzhōu Chinese is typically spoken, while Standard Chinese predominates in tasks involving reading and writing. As Sūzhōu Chinese and Standard Chinese can be written with the same hànzì, giving a Sūzhōu Chinese reading for each stimulus character requires some additional mental work for most participants, and familiarizing participants with the stimuli was non-trivial. Prior to experimental procedures, participants began by reading each stimulus item aloud without any guidance from the experimenter; they were prompted to try different (unspecified) readings if their default reading for a given character was borrowed from Standard Chinese (e.g. 包 'bag' read with Standard Chinese segments as [pɑʊ^{44}], rather than expected Suzhou [pæ^{44}]) or otherwise significantly diverged, most often due to incorrect application of a phonological correspondence between Sūzhōu Chinese and Standard Chinese.

As a result of the nature of the reading task, unexpected readings were occasionally assigned to stimuli, both as Standard Chinese readings and readings that were appropriate as a Sūzhōu Chinese reading but unanticipated in the experimental design for the test item in

---

[4] The bilabial and labiodental fricative vowels, which are allophones of /əu/ that follow bilabial and labiodental onset consonants, respectively, are discussed further in Chapter 3. They exhibit visible constrictions of the types described; the bilabial vowel has loose compression of the lips that occasionally results in bilabial trilling.

		Readings		
	Item	Expected	Major variants	Discard rate
Ø onset	衣 ‘clothes’	ɪ͡ʑ^{44}	—	1 (0.22%)
	烟 ‘smoke’	i^{44}	—	4 (0.88%)
	迂 ‘circuitous’	y͡ʑ^{44}	—	2 (0.45%)
	怨 ‘blame’	y^{523}	—	3 (0.66%)
	凹 ‘concave’	æ^{44}	—	**5 (1.1%)**
	哑 ‘mute’	u^{51}	—	0
	蛙 ‘frog’	u^{44}	wa^{44}	**37 (12.98%)**
/ɕ/ onset	希 ‘rare’	ɕɪ͡ʑ^{44}	—	3 (0.67%)
	掀 ‘flip’	ɕi^{44}	ɕjɛ^{44}	2 (0.44%)
	虚 ‘weak’	ɕy͡ʑ^{44}	—	1 (0.22%)
	休 ‘rest’	ɕy^{44}	seɪ^{44}	**5 (1.10%)**
	箫 ‘flute’	*ɕjæ^{44}	ɕjɛ^{44}, sjɛ^{44}	0
	靴 ‘boot’	*ɕu^{44}	ɕy^{44}, ɕɥəʔ^{5}	3 (0.67%)
/s/ onset	西 ‘west’	sɪ͡ʑ^{44}	ɕɪ͡ʑ^{44}	1 (0.22%)
	C ‘C’	sɪ͡ʑ^{44}	ɕɪ͡ʑ^{44}	0
	鲜 ‘fresh’	si^{44}	ɕi^{44}	2 (0.45%)
	修 ‘repair’	*sy^{44}	ɕy^{44}, seɪ^{44}	1 (0.22%)
	烧 ‘roast’	sæ^{44}	—	1 (0.22%)
	沙 ‘sand’	su^{44}	—	**5 (1.12%)**
/p/ onset	比 ‘compare’	pɪ͡ʑ^{51}	—	2 (0.45%)
	边 ‘side’	pi^{44}	—	1 (0.22%)
	包 ‘package’	pæ^{44}	—	0
	疤 ‘scar’	pu^{44}	—	**8 (1.77%)**
Other	丝 ‘thread’	sɿ^{44}	—	2 (0.45%)
	书 ‘book’	sʮ^{44}	—	1 (0.22%)
	夫 ‘husband’	fə͡v^{44}	—	1 (0.22%)
	播 ‘spread’	bə͡β^{44}	boʔ^{5}	**21 (4.65%)**
	儿 ‘child’	*əl^{523}	ȵɪ͡ʑ^{523}, ɑɻ^{523}	1 (0.22%)

Table 4.2: Stimuli by expected reading, from Xing (2014). An asterisk next to an expected reading indicates that no readings had the value expected from Xing (2014). Shading indicates that onset /s/ in the stimulus was affected by palatalization.

question. Standard Chinese readings were occasionally produced, mainly for the character 蛙 “frog”, whose intended reading [u^{44}] contains no segments in common with the Standard Chinese reading [wɑ^{44}]. A smaller portion of all other stimuli with an intended Sūzhōu Chinese reading containing [u] were similarly misproduced with an [ɑ]. All other fluent readings containing [ɑ] were discarded. A handful of other productions showed some characteristics of both the canonical Standard Chinese reading and the canonical Sūzhōu Chinese reading, for example, assigning the test item 烧 a reading of [sɑʊ^{44}] (which uses the rhyme but not the onset of Standard Chinese [ʂɑʊ^{55}]). These were also discarded.

Some stimuli were subject to variable readings even among fluent Sūzhōu Chinese speakers, primarily as a result of sound changes in progress. The result was a somewhat unpredictable token count for some of the target segments. A small number of misproductions were due to single confusions among stimulus characters or were otherwise idiosyncratic, affecting single productions. Otherwise, the biggest systematic contributor to this variation is palatalization of /s/, which may affect any /s/-initial reading that is followed by /i/ or /i͡z̩/, and which results in the /s/ initial being produced as [ɕ]. These stimuli are indicated with shaded cells in Table 4.2. Several speakers produced two stimuli, 播 ‘spread’ and 靴 ‘boot’, with checked rhymes (closed with a glottal stop) instead of the expected open-syllable rhymes; the resulting readings have properties distinct from the intended reading that make them ill-suited for analysis and were discarded. Four stimuli never exhibited the reading expected from source material, but the variant pronunciations that did occur still contained some desired segments (particularly the onset consonants /ɕ/ and /s/). All desired segments from these variant readings are retained, since regardless of reading, they do not clearly represent an excursion into a Standard Chinese reading style.

Additions or replacements to the set of stimuli were undertaken for two items in an attempt to reduce variability. The letter “C”[5] was added to the stimulus set as a hedge against variability in palatalization of initial /s/, but was discontinued after two participants since it was discovered to be subject to the same variable palatalization as existing stimulus items. As mentioned above, the stimulus item intended to be read as [u^{44}], “frog”, was rather badly affected by misreadings as [wɑ^{44}]. Starting with Subject 28 (out of 44), this character was replaced with the more familiar character 哑 “mute”, which was read nearly consistently as the intended Suzhou Chinese reading [u^{51}].[6]

Other factors affected the number of stimuli recorded by each participant. Occasional errors in the timing of key presses resulted in acquisitions with no usable target material. On a handful of occasions, participants also repeated desired target material multiple times in a single frame sentence due to an ambiguity in the experiment’s description text; in

---

[5] Readers of Chinese regularly encounter Roman letters, mainly in alphanumeric codes (seat numbers, service tickets, etc.); familiarity with reading and interpreting them is very general.

[6] It may come as a surprise that the desired reading of 哑 “mute” as [u^{51}] is markedly more familiar than the desired reading for 蛙 “frog” as [u^{44}]. I speculate that this has to do with the local popularity of a restaurant called 哑巴生煎 [u^{51} tsɿ^{35} sã^{44} tɕi^{21}] or [u^{51} pu^{35} sã^{44} tɕi^{21}] (literally “mute fried dumplings”), apparently named for patrons’ awestruck reaction to the namesake dish. Many participants who reviewed the desired reading of 哑 immediately recognized the character as a part of the restaurant’s name.

this case, the extra tokens were all marked for analysis unless some obvious disfluency had occurred. Participants were also asked to record two to three additional blocks at the end of the experiment if there was a particularly high rate of misreading or timing errors evident while the experiment was being run.

The total number of acquisitions discarded for timing and misreading errors of all sorts was 106 across all participants, with the most errors committed by any one participant being 12 (Subject 20). The remaining trials contain a total of 11,613 productions of target syllables across all participants. Misreading and timing error rates generally do not exceed 1% for any one stimulus pooled across all participants. Stimuli read with an /u/ are misread and discarded somewhat more often than other characters, in excess of 1%, in particular "frog", which was misread and discarded nearly 13% of the time. The total number of tokens across all stimuli collected for a given participant thus ranges from 338 to 209, with a median value of 260.[7] A breakdown of each participant's readings of each stimulus item, and the number of tokens of target segments for each participant, are given in Appendix C.

4.2.4 Frame sentence

All stimuli were produced in the following frame sentence:

我看到 _____ 该个字哉。
[ŋəu^{24} kʰø^{51} tæ^{35} _____ kɛ^{44} kəʔ^{5} zɿ^{33} tsɛ^{21}]
"I see _____ , that character."

Subjects were instructed to produce this specific reading of the sentence, with focus given to maintaining a consistent segmental context surrounding the target. However, given that the frame sentence was presented in hànzì, some small amount of variation in the reading of the frame sentence is inevitable. Variant frame sentence readings affecting the segmental material immediately surrounding the target (i.e., the reading of the preceding character 到 and the following character 该) occurred in a small number of acquisitions overall and as such were not expected to present a major confound for analysis. Nonetheless, some details follow.

Variant frame sentence readings can be broken into two types, those affecting the content immediately preceding the target word and those affecting the immediately following content. In a few instances, speakers replaced the character immediately preceding the target word, 到 (intended reading [tæ^{35}]), with a reading [tɕi^{35}] more commonly associated with the semantically related character 见.[8] This replacement affected only 23 acquisitions out of the 11,616 collected. The most [tɕi^{35}] readings employed by any one participant was 15 (Speaker 25), with a median of 0 across all participants. An additional 64 acquisitions

---

[7] Speaker 19, the speaker with the token count of 209, was unable to run ten blocks of the study due to a technical mishap, instead running only eight.

[8] Both 看到 and 看见 have the sense of "see, perceive"; it is not clear which is more commonly in use in Sūzhōu Chinese.

had a disfluency that resulted in something other than the frame sentence preceding the target, most commonly a partial production of the target preceding a full production (e.g. [s .. sæ^{44}]). Except in cases where the target was very clearly produced twice or the first production was a licit Sūzhōu Chinese production containing desired segments, only the second repetition was analyzed.

Speakers occasionally read the character immediately after the target, 该 [kɛ^{44}], as [ʔeɪ^{44}]; these variants turn out to be a salient sociolinguistic variable which places individual speakers in different areas within the Northern Wú region (Wang Feifan, pers. comm.). Participants were instructed to use the [kɛ^{44}] variant as consistently as they could, but a number of productions are followed by the [ʔeɪ^{44}] variant instead. One participant (Speaker 3) exclusively used the [ʔeɪ^{44}] reading (in all 336 of her productions) except for two [kɛ^{44}] readings. Excluding this participant, only 54 acquisitions used the [ʔeɪ^{44}] reading out of a total of 10,750 [kɛ^{44}]. The most readings of [ʔeɪ^{44}] produced by any one participant was 14 (by Speaker 8), with a median value of 1 across all participants. All articulatory analyses are within-speaker, so a participant's choice of [ʔeɪ^{44}] or [kɛ^{44}] readings for 该 should not be a concern, so long as the participant was consistent in choosing this reading of the frame sentence. Disfluencies and repetitions immediately following a target reading totaled 40 out of the 11,616 total productions.

4.2.5 Procedure

Participants were familiarized with the stimuli and frame sentence before making the recording, as described above. Following introduction to the materials, the stabilization helmet was fit to the participant's head and lighting was adjusted for purposes of capturing face video. The equipment was subsequently powered on and its connections checked with two practice acquisitions. Following this, and after participants had been instructed on how to complete a trial using the experiment software, the experiment proper began. Each experiment run consisted of ten blocks. In each block the 26 stimuli were presented in random order in the frame sentence. Trials were self-paced, with the beginning and ending of ultrasound and audio recording triggered by keypresses. At the end of the final block, I assisted the participant in removing the stabilization helmet, compensated them for their time and effort, and asked them to continue the snowball sample by forwarding other potential participants.

All study procedures were conducted in spoken and written Standard Chinese, both on the consent and survey forms and in the experiment's instructions themselves. On occasion, younger participants spoke English with the experimenter, primarily while trying to interpret the on-screen instructions and during helmet application or removal.

4.3 Processing and analysis

Acoustic data from all participants, and articulatory data from all but one participant, were analyzed as described in this section. Speaker 35 exhibited an unstable or nonexistent contrast between /i/ and /i͡z̞/ on the one hand and no contrast between /y/ and /y͡z̞/ on the other hand. This speaker produced no tokens that could reasonably be transcribed as [y͡z̞], and roughly half as many tokens of [i͡z̞] ($n = 20$) as was typical for other participants. Participant 35's data are thus included in the acoustic analyses in Section 4.4.1 but are altogether excluded from the articulatory analyses in Sections 4.4.2–4.4.3, given that the dimensionality reduction process described in Section 4.3.3 yields a less informative representation of a speaker's activity if less data is used as a basis.

Only certain target segments from recorded productions of test items were used for analysis. All tokens of /s/, /i/, /ɕ/, /i͡z̞/, /ɿ/, and /y͡z̞/, including the latter's allophone [ɥ], contained in stimulus productions are used for articulatory analysis in some way. All of the vowels in this set, plus all tokens of /u/, /æ/, and the two labial fricative vowels [ə͡v] and [ə͡β], are used for description of the acoustic space. Some infrequent readings of stimuli, such as the reading of 播 "spread" as [boʔ^{5}], contain no target segments, and are discarded for this reason. All productions of 儿 "child" were removed from the set used for analysis, given that the readings given either contained no target segments, such as [ɑɻ^{523}], or had a nasal initial, such as [ȵi͡z̞^{523}] or [ȵi^{523}]; the coarticulatory effects of a nasal initial on the immediately following vowel are not clear. Occasional readings of 怨 "blame" as [ȵy^{523}], rather than expected [y^{523}], are also discarded for the same reason. Thus, of the 11,613 productions collected, only the 11,136 productions remaining after these are actually analyzed acoustically, and only 8,056 contain tokens that are analyzed articulatorily.

4.3.1 Transcription

The audio data obtained were submitted to forced alignment using the Penn Forced Aligner (P2FA) (Yuan and Liberman, 2008). Since P2FA is trained on English and produces output in English phones, and Suzhou Chinese has a number of typologically unusual vowels that English lacks, a mapping between English phone transcription and the actual phonetic content of the recordings was used. This was achieved by substituting an English phone with similar acoustic priors to the actual Suzhou target segment. The full set of substitutions is shown in Table 4.3; identical forced aligner labels applied to contrastive phones were disambiguated by word label. The forced aligner performed well at aligning these labels with the Suzhou segments, but to ensure consistency, the TextGrids were inspected and alignments manually corrected in Praat (Boersma and Weenink, 2017), with special attention paid to the vowels that do not have a close acoustic match in any English vowel.

Transcriptions used for forced alignment were, for the most part, automatically generated from the expected pronunciation of the target word combined with the frame. However, occasional manual interventions were required, mainly due to inter- and intra-subject variation in pronunciation of target or frame segments. For instance, most subjects produced the frame sentence character 该 as [kɛ^{44}] consistently, with some speakers varying between [kɛ^{44}] and [ʔeɪ^{44}] from acquisition to acquisition, necessitating a change to the frame sentence's segments. Speakers' pronunciations of target words also varied, as discussed above in Section 4.2.3. Any and all pronunciation variants containing target segments and produced

Sūzhōu Chinese phone(s)	P2FA label
/i/, /y/, /i͡ʑ/, /y͡ʑ/	IY1
/ɿ/, [ʮ]	IH1
/u/	UW1
[ə͡v], [ə͡ß]	UH1
/æ/	AE1
/s/	S
/ɕ/	SH

Table 4.3: Transcriptions for target segments in target Suzhou Chinese words, as supplied to the Penn Forced Aligner (Yuan and Liberman, 2008).

without disfluencies were force-aligned and included in the set of segments used for analysis; that is, the segments in an unexpectedly palatalized token of 西 [si͡ʑ^{44}] ‘west’, produced as [ɕi͡ʑ^{44}], are included as tokens of /ɕ/ and /i͡ʑ/ in the analysis.

4.3.2 Acoustic analysis

Two acoustic analyses were used to characterize the collected vowel data. Fundamental frequency (F0) and formants (F1–F3) were extracted using an inverse filter control method (Ueda et al., 2007) at seven evenly spaced time points over the duration of each vowel, with the first time point at the onset of the vowel and the seventh point at the offset of the vowel. Once extracted, formants were log-mean normalized to remove physiological factors from the data while preserving linguistically relevant information in formant values (Nearey 1977). Harmonic-to-noise ratio (HNR), or the ratio of the spectrum’s periodic and aperiodic components, was calculated in Praat (Boersma and Weenink, 2017) using the algorithm described in Boersma (1993). HNR data was z-scored to normalize for inter-subject physiological differences and differences in the amount of aperiodic noise in the environment during the various recording sessions.

4.3.3 Ultrasound processing

As alluded to above, the synchronization pulse train generated by the ultrasound device was recorded as an additional channel to the recorded audio. Using the synchronization pulse train signal, single ultrasound frames were extracted from the captured raw binary data at locations in the data packet corresponding to the acoustic midpoint of the target vowel or fricative. Empty or incorrectly sized data were discarded, affecting a total of 5 ultrasound recordings across the 11,613 made for all subjects. Frames were then converted to the characteristic ultrasound “fan” shape, to physical scale, from the compressed rectangular format of the raw binary data. Fan-converted frames were preprocessed using a median filter with a three-pixel radius as an easily-implemented means of reducing the speckle noise characteristic of ultrasound imaging (see Kak and Rosenfeld, 1982, Ko and Lee, 1991). No further preprocessing alterations, such as a region-of-interest mask, were applied to the data.

Ultrasound image data has the twin disadvantages of being extremely high-dimensional—here, each extracted frame's 24,198 pixels can range in value from 0 (black) to 255 (white)—and relatively noisy in spite of measures to reduce noise. To improve the data in both aspects, relevant frames (defined below) were entered into a principal components analysis (PCA) (see Hueber et al. (2007) and Mielke et al. (2016) for similar approaches). A separate PCA was run for each subject in this way; the number of PCs output ranges from about 20 to 40, representing a reduction in dimensionality of several orders of magnitude. The output of PCA for each subject was used to train two linear discriminant analyses (LDAs) that are described further in Sections 4.4.2 and 4.4.3 below.

4.4 Results

Results are discussed below in turn for the articulatory data (Sections 4.4.2 and 4.4.3) and the acoustic data, in a more preliminary fashion (Section 4.4.1). The two analyses undertaken on the articulatory data address, respectively, Hypotheses 1a and 1b, on whether speakers show uniformity among frication-producing sounds at the lamino-postalveolar place of articulation, and Hypothesis 2, on whether fewer and more uniform articulatory strategies are employed for /i͡ʑ/ and /y͡ʑ/ at the population level as apparent time passes. The articulatory findings confirm Hypotheses 1a and 1b for a large majority of study participants, suggesting a general role of articulatory uniformity in determining how a speaker's inventory of articulatory strategies for fricative vowels and fricative consonants are organized. I reject Hypothesis 2, however, based on a complete absence of evidence for change in apparent time in this larger sample of speakers compared to the four speakers in the pilot study.

On the basis of classification of the articulatory data by a linear discriminant analysis, I find that speakers overwhelmingly produce their postalveolar fricative vowels in ways closely associated with their individual strategies for producing the alveolopalatal fricative /ɕ/; the same is true for their apico-alveolar vowels and their apico-alveolar fricative /s/. Crucially, for many speakers, the strategy used for /i͡ʑ/ and /y͡ʑ/ appears to be relatively invariant with respect to segmental context, most notably the presence or absence of an immediately preceding, tautosyllabic /ɕ/ onset.

4.4.1 Preliminary acoustic characterization of the fricative vowels

By-subject means for the first and second formant frequencies of each of the target vowels are shown in Figure 4.4. Several characteristics of the fricative vowels merit some discussion in light of the articulatory data to follow. Dorso-palatal /i/ and /æ/ are acoustically well separated from the other vowels. In particular, there is generally a large acoustic gap between

/i/ and the lamino-postalveolar vowel /i͡ʑ/. In contrast, the rest of the vowels overlap considerably in the area of acoustic space most commonly associated with central and back high vowels (e.g., [ɨ], [u]).

Two subsets of these vowels occupy nearly the same location in the aggregate: dorso-palatal /y/ and the lamino-postalveolar vowels /i͡ʑ/ and /y͡ʑ/ in one group, and the two apico-alveolar vowels /ɿ/ and [ʮ] in the other group. Considering F3 as a potentially distinguishing characteristic (Figure 4.4, bottom) only separates out /i͡ʑ/ from /y͡ʑ/ and /y/, which themselves remain overlapping in F1-F2-F3 space; the two apico-alveolar vowels similarly overlap. The overall placement of these groups, and the identity of the overlapping vowels in each group, is the same as that found in Ling, (2009, pp. 18, 24–25).

In terms of their location in F1-F2 space relative to high front /i/ and high back /u/, the lamino-postalveolar vowels in Sūzhōu Chinese have a clear resemblance to fricative vowels that have been described acoustically in other languages. In particular, there is a strong resemblance between the lamino-postalveolar vowels /i͡ʑ/, /y͡ʑ/ and the “Viby-colored” vowels in regional Swedish, particularly as described in Westerberg (2016) and Schötz et al. (2011). Connell (2007) also describes the lamino-postalveolar vowel of Len Mambila, which he transcribed as [ʒ͡ɨ], as having a “high centralized” quality, in this case seemingly indicating an F2 lower than that of the language’s /i/. The apico-alveolar vowels in Sūzhōu Chinese /ɿ/, [ʮ] have a markedly lower F2 compared to even the lamino-postalveolar vowels in these descriptions, which is not described for any vowel in, e.g., Swedish or Len Mambila, but falls into the same range as the Standard Chinese “apical” vowels (Zee and Lee, 2007, Ling, 2009, pp. 23–25).

The harmonic-to-noise ratio (HNR) of the six target vowels (Figure 4.5) reveals some additional acoustic distinctions. In the aggregate, there is a clear difference in HNR between each lamino-postalveolar and dorso-palatal pair (/i/, /i͡ʑ/ and /y/, /y͡ʑ/). In both cases, the lamino-postalveolar vowel typically has a lower HNR than the dorso-palatal vowel, indicating the presence of a greater proportion of aperiodic energy in the vowel spectrum. Apico-alveolar vowels, interestingly, are typically somewhere between the other two groups in terms of HNR levels: this is somewhat unexpected given that they obligatorily occur following an apico-alveolar fricative or affricate, which would produce aerodynamic conditions conducive to a relatively high level of aperiodic energy on the following vowel.

The HNR contrast is less consistent for the rounded pair /y/-/y͡ʑ/. This is in part due to the fact that /y/ has a bimodal distribution in HNR (Figure 4.6). This leaves it unclear how some speakers distinguish these vowels on the basis of acoustic characteristics alone. It is in fact possible, from this data, that some speakers have a single merged category for /y/ and /y͡ʑ/. This largely replicates findings in Ling (2009), although Ling’s results do not hint at the variability and potential loss of discriminative power of /y/’s HNR compared to that of /i/ (pp. 22–23).

Duration also differs among the six vowels discussed here: the two apico-alveolar vowels /ɿ/ and [ʮ] show a marked tendency to be shorter than the other vowels, as shown in Figure 4.7. Rose (1982) observes a similar pattern in Zhènhǎi 镇海 Chinese, a close relative of Sūzhōu Chinese, in that the apico-alveolar vowels have a strikingly short duration compared

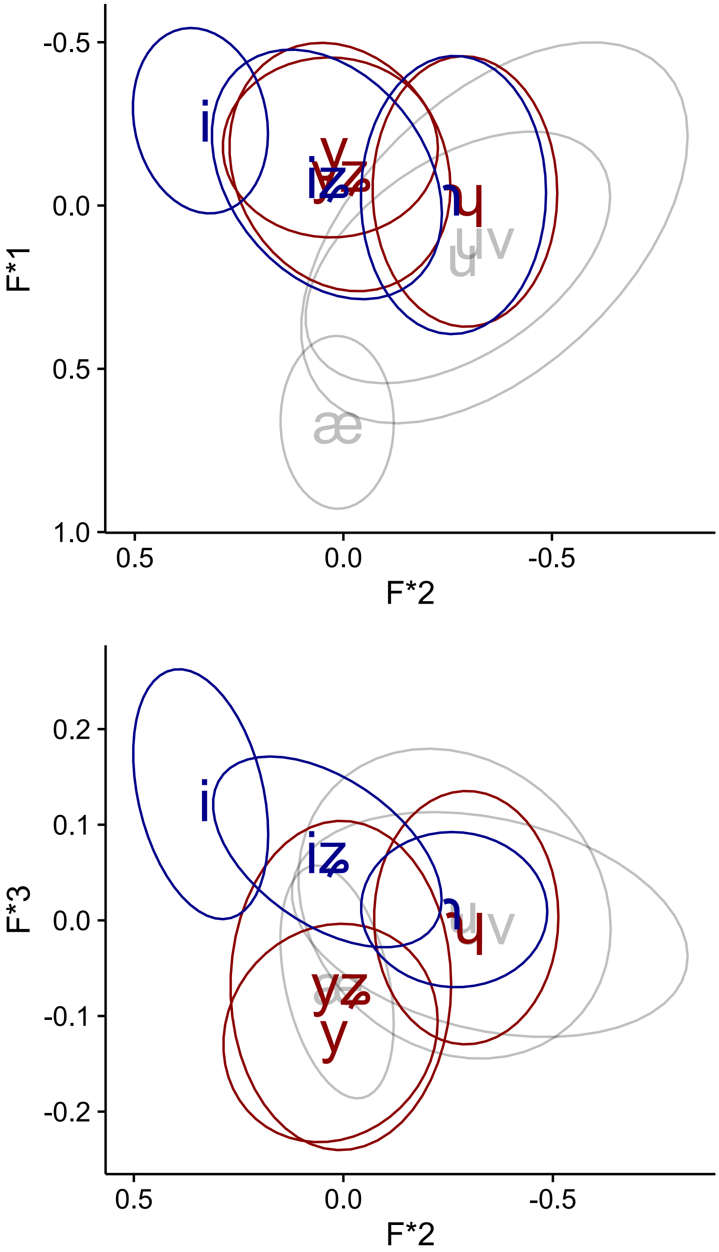


Figure 4.4: By-vowel means (pooled across all participants) of Nearey-normalized F1, F2, and F3 in target vowels, with 95% confidence ellipses. Red ellipses and symbols are for the rounded vowels [ʮ], /y͡ʑ/, /y/; blue ellipses and symbols are for the unrounded vowels /ɿ/, /i͡ʑ/, /i/. Light gray ellipses mark /u/, the [ə͡v] allophone of /əu/, and /æ/.

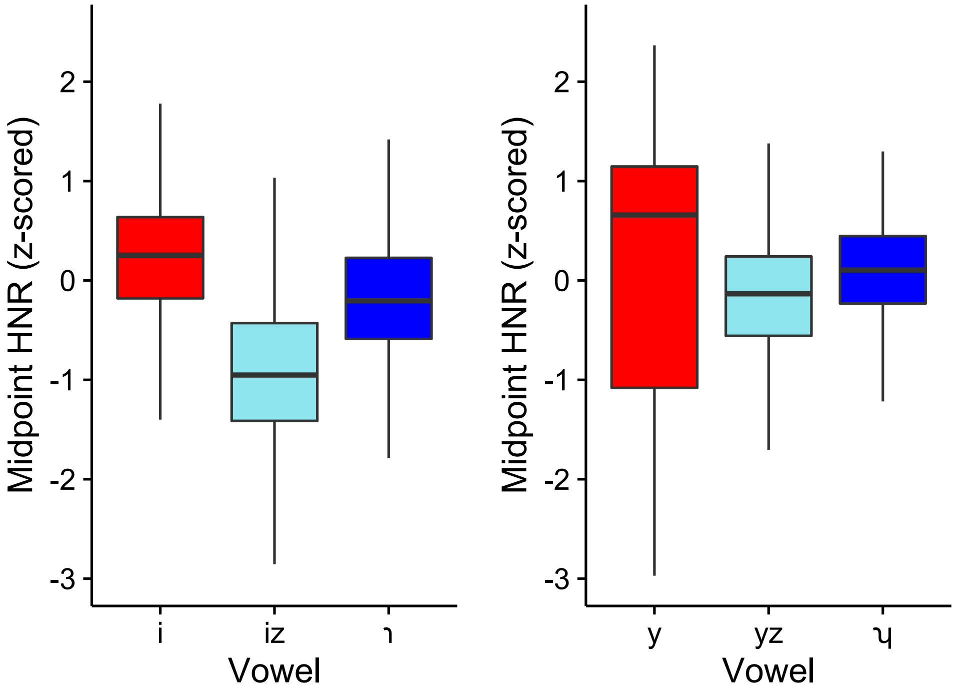


Figure 4.5: Z-scored harmonic-to-noise ratio pooled across all speakers by vowel, in unrounded (left) and rounded (right) groups.

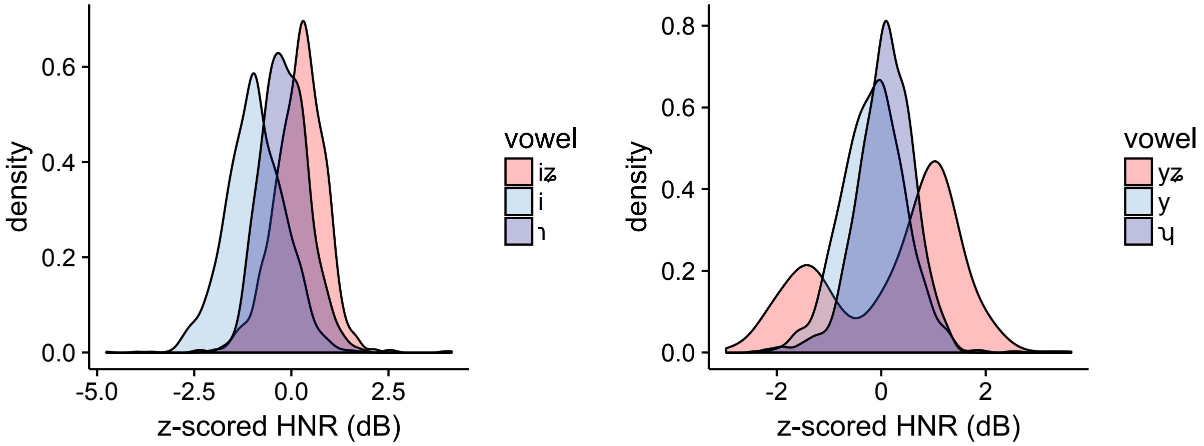


Figure 4.6: Probability density of HNR at acoustic midpoint of vowel, for unrounded (left) and rounded (right) series of target vowels, pooled across all participants.

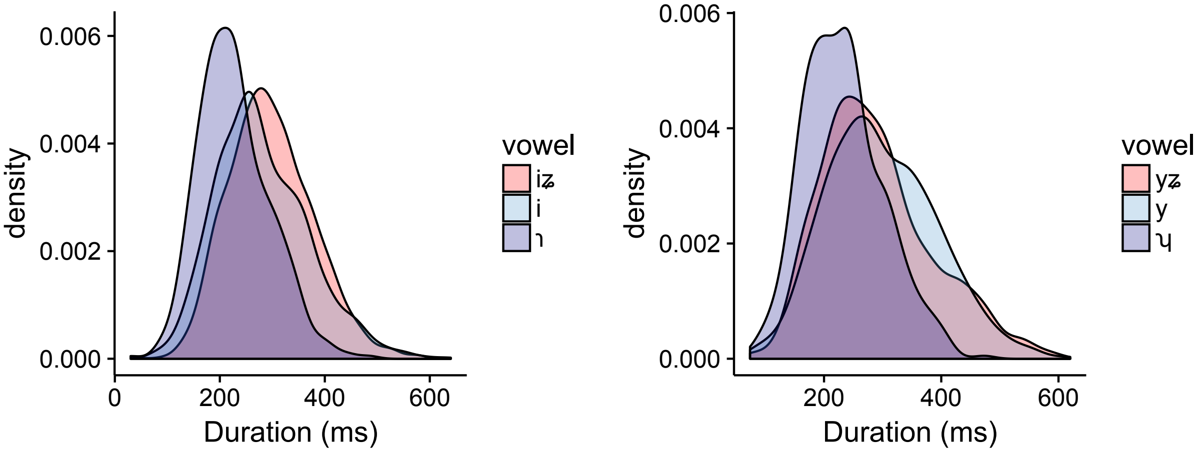


Figure 4.7: Probability density of duration measured for vowel, for unrounded (left) and rounded (right) series of target vowels, pooled across all participants.

to other vowels when placed in the same consonantal context. As in Sūzhōu Chinese, the apico-alveolar vowels are obligatorily preceded by an apico-alveolar fricative or affricate, e.g. /s/; in Zhènhǎi Chinese this fricative is also described as lengthened relative to tokens of /s/ that precede other vowels.

4.4.2 Articulatory characteristics of the fricative vowels

This section reports the results of a linear discriminant analysis (LDA) on the ultrasound data to test whether the tongue shapes typical of the fricative vowel pairs /i͡ʑ/, /y͡ʑ/ and /ɿ/, [ʮ] are most similar to that of the fricatives /s/ and /ɕ/ or the vowel /i/. This analysis suggests that the articulations of the Sūzhōu Chinese apico-alveolar and lamino-postalveolar vowels closely resemble a given speaker's fricative consonants /s/ and /ɕ/, respectively, with some speakers showing a closer overlap of vowel and fricative categories than others. On the basis of the review of uniformity pressures in Section 4.4.1, and given the considerable similarity (in descriptive terms) between Sūzhōu Chinese's consonantal fricatives and its fricative vowels, it is an empirical question whether these articulatory strategies are merely very similar or in fact identical (and based off of the same basic motor program, subjected to some minor adjustments).

Examining ultrasound images of the tongue (Figure 4.8) suggests that Sūzhōu Chinese does generally exhibit a three-way contrast in vowel constriction location as described in Ling (2009): between rounded and unrounded vocoids with (1) apico-alveolar (/ɿ/, [ʮ]), (2) lamino-postalveolar (/i͡ʑ/, /y͡ʑ/), and (3) dorso-palatal (/i/, /y/) constriction locations, all of which have characteristic tongue postures that are clearly visible in ultrasound data. To examine this tendency across the entire data set, ultrasound frames were extracted from longer video files at the acoustic midpoints of target segments, as discussed above.

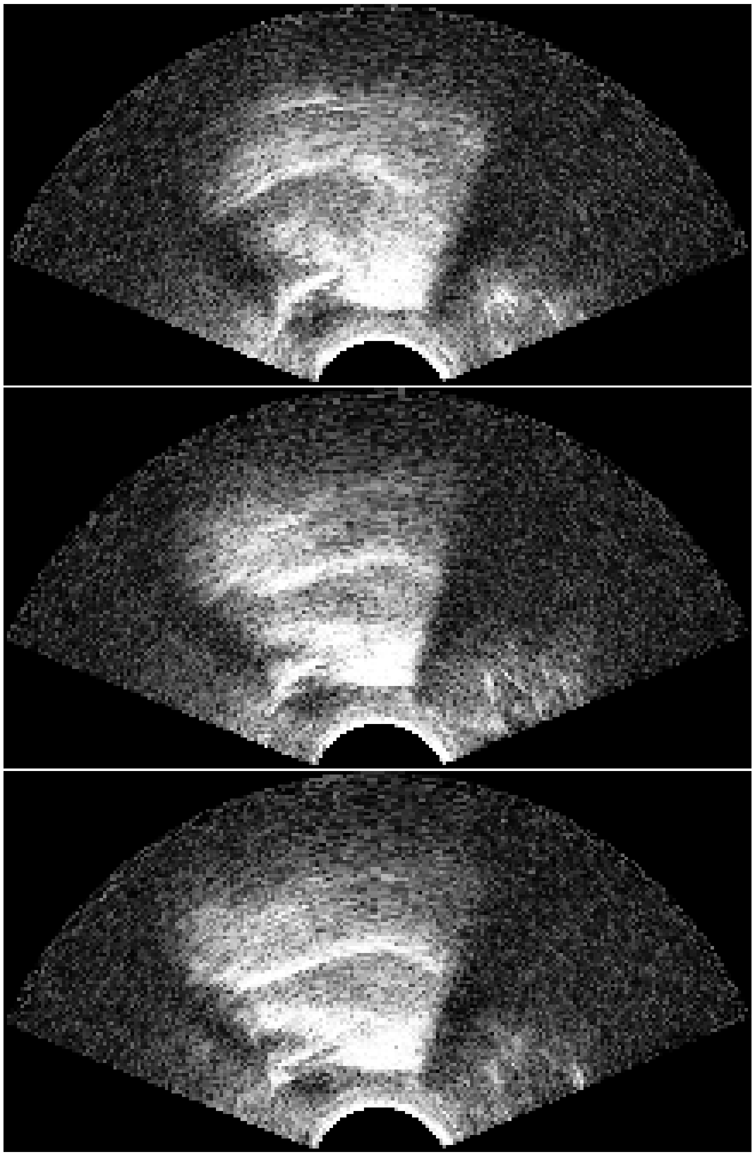


Figure 4.8: Ultrasound images of the tongue for Speaker 3 at the acoustic midpoint of the vowels in the minimal triplet (from top to bottom) 鲜 [si^{44}] ‘fresh’, 西 [si͡z̞^{44}] ‘west’, and 丝 [sɿ^{44}] ‘thread’. Left is anterior. Note different tongue blade positions for [i͡z̞] and [ɿ].

The set of midpoint frames from each subject were submitted to a principal components analysis (PCA) to reduce dimensionality, along the same lines as Hueber et al. (2007) and Mielke et al. (2016). The lesser between 20 principal components (PCs) or PCs sufficient to account for 50% of variation in the data were included in the transformed data, resulting in between 20 and 40 PCs per participant. Since any ultrasound frame in the basis data can be reconstructed through rescaling and combining the PCs, PCs derived from tongue ultrasound imaging data are sometimes referred to as "eigentongues" (Hueber et al., 2007). To illustrate the nature of PCA's outputs, minimum and maximum loadings of the first two PCs (PC1 and PC2) for Speaker 3 are shown in Figure 4.9. Bright and dark areas in each plot are light-colored in the basis data for extremely high and low loadings, respectively, of the appropriate principal component. The principal components depicted in the loading plots encompass the range of variation between these extremes.

The basis data were then separated into two groups, the training set and the test set, in an effort to characterize the latter group's articulatory strategies using the former group's. The three segments in the training set—apico-alveolar /s/, lamino-postalveolar /ɕ/, and dorso-palatal /i/—have known articulatory properties and are plausible models for characterizing the articulatory strategies of the test set: the fricative vowels /ɿ/, [ʮ], /i͡ʑ/, /y͡ʑ/. To carry out these comparisons, the principal component scores, or the rescaling factors applied to each principal component's "eigentongue" to reproduce a given item in the basis data, were used as a low-dimensional input to a linear discriminant analysis (LDA). PC scores were used to train an LDA to distinguish the three training segments, characterizing them in a two-discriminant space defined by new scores on linear discriminants 1 and 2 (LD1 and LD2). PC scores for the test segments were then transformed into the resulting LD1-LD2 space, yielding LD1 and LD2 scores and a classification judgment (as either /i/, /ɕ/, or /s/) for each test vowel token on the basis of the obtained scores. The classification of test segments resulting from the LDA (Figure 4.10) shows that the fricative vowels typically described as apico-alveolar (/ɿ/, [ʮ]) are nearly always classified as /s/, and the fricative vowels described as lamino- or dorso-postalveolar (/i͡ʑ/, /y͡ʑ/) are generally classified as /ɕ/, suggesting that the lingual postures of these vowels more closely resemble these consonants than /i/.

Median LD1 and LD2 values for twelve speakers' training and test segments are provided in Figures 4.11 and 4.12 to illustrate some gradient patterns in the data set not readily apparent in the classification data. Most speakers produce the apico-alveolar vowels /ɿ/ and [ʮ] virtually identically to /s/, with some subjects exhibiting articulatory strategies that differ in idiosyncratic ways, most often in the direction of /i/ in LD1-LD2 space. Unlike the postalveolar vowels, the apico-alveolar vowels obligatorily follow an apico-alveolar fricative or affricate, meaning that all tokens of /ɿ/ and [ʮ] used in this study have significant exposure to coarticulatory influence from /s/. The general close similarity of /s/ and the alveolar vowels /ɿ/ and [ʮ] is thus a useful check on the informativeness of the LDA, given that this close similarity is expected for all speakers examined here, based on prior articulatory phonetic studies of these "apical vowels" in Sūzhōu Chinese (Ling, 2009) and other Chinese varieties (Lee-Kim, 2014; Faytak and Lin, 2015).

Unlike the apico-alveolar vowels, the postalveolar vowels /i͡ʑ/, /y͡ʑ/ do not always co-

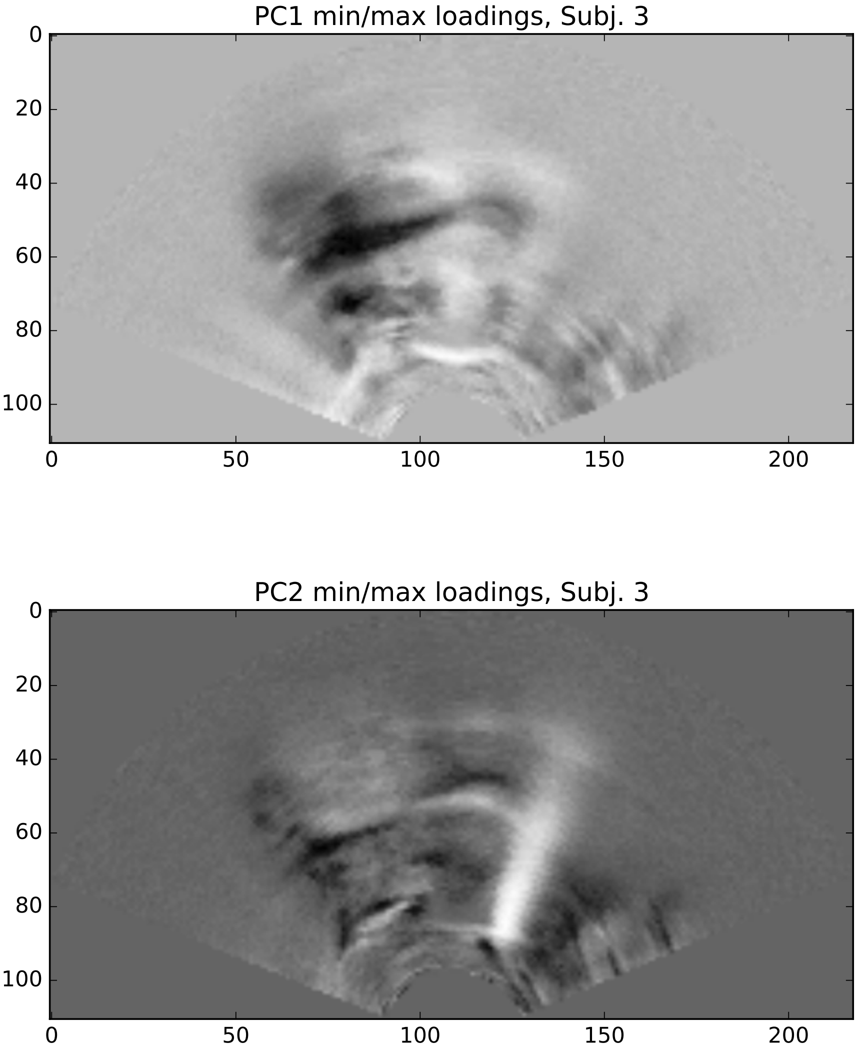


Figure 4.9: Loadings for Speaker 3's first two principal components (PC1–PC2) mapped to the original shape of the ultrasound image data. Note the /i/-like (bright) and /s/-like (dark) tongue contours visible in PC1; compare Figure 4.8. Left is anterior.

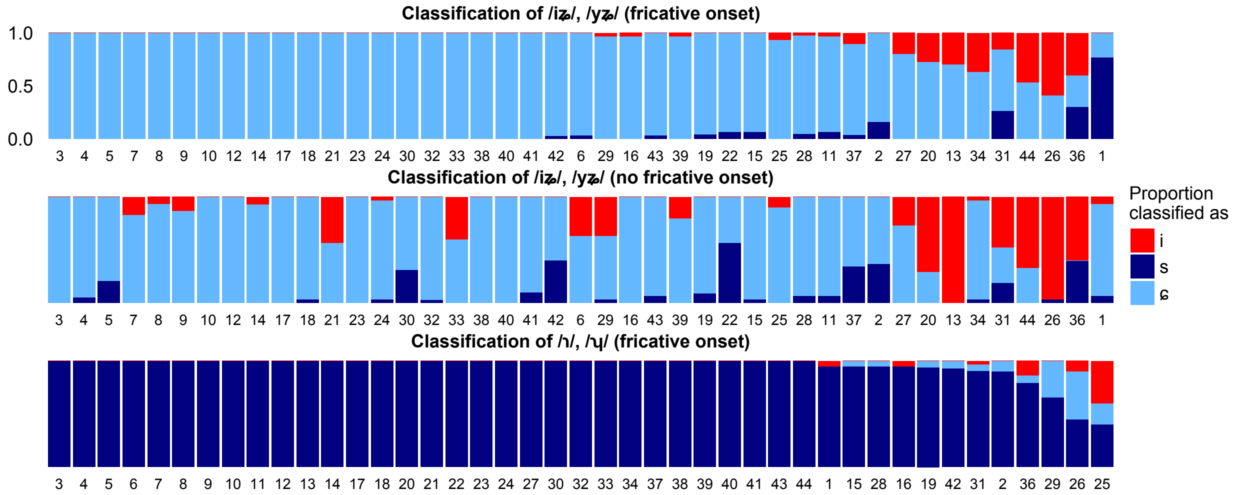


Figure 4.10: LDA classification of postalveolar and apico-alveolar test items. Subject results are sorted by proportion classified as /ɕ/ in the fricative onset condition.

occur with a homorganic fricative onset, with roughly half of the tokens used in this study having with no onset or a /p/ onset. Unsurprisingly, then, patterns of articulation for the vowels /i͡ʑ/, /y͡ʑ/ relative to the consonant /ɕ/ are more varied. The participants whose LD1-LD2 spaces are shown in Figure 4.11 comprise a group who illustrate uniformity in their articulation of the vowels /i͡ʑ/, /y͡ʑ/ and the consonant /ɕ/. The participants in Figure 4.12, on the other hand, appear to have an articulatory strategy for /i͡ʑ/ and /y͡ʑ/ distinct from both /i/ and /ɕ/, most often intermediate between /i/ and /ɕ/ (and not, i.e., /i/ and /s/ or /ɕ/ and /s/). In this latter group, an effect of fricative onset appears to be present in terms of relative similarity of /i͡ʑ/, /y͡ʑ/ to /i/: when a fricative onset is absent, /i͡ʑ/ and /y͡ʑ/ are produced more similarly to /i/, although this effect varies idiosyncratically. This degree of variation is unexpected from prior articulatory data reported in Ling (2009), but is consistent with my pilot data as discussed in Section 4.1.1: the two groups of speakers in Figures 4.11 and 4.12 appear to exhibit patterns of articulatory behavior analogous to pilot Speakers 01 and 08, respectively.

A close examination of two speakers typical of the uniform and non-uniform groups aids in relating the classification data to actual articulatory strategies in LD1-LD2 space. Figure 4.13 shows individual tokens in LD1-LD2 space and each token's classification results. If all of a vowel's productions are clustered tightly about one of the training segments in LD1-LD2 space, as is the case for all test vowels for Speaker 3 (Figure 4.13, left), classification is overwhelmingly as that training segment. In this case, Speaker 3's /i͡ʑ/ and /y͡ʑ/ are consistently classified as /ɕ/ due to this distribution. Other speakers' /i͡ʑ/ and /y͡ʑ/ productions occupy a location intermediate between two training categories, as is the case for Subject 13 (Figure 4.13, right). Speaker 13's /i͡ʑ/ and /y͡ʑ/ consequently straddle the classification function, with the resulting mixed classification between /i/ and /ɕ/ clearly visible in Figure 4.10.

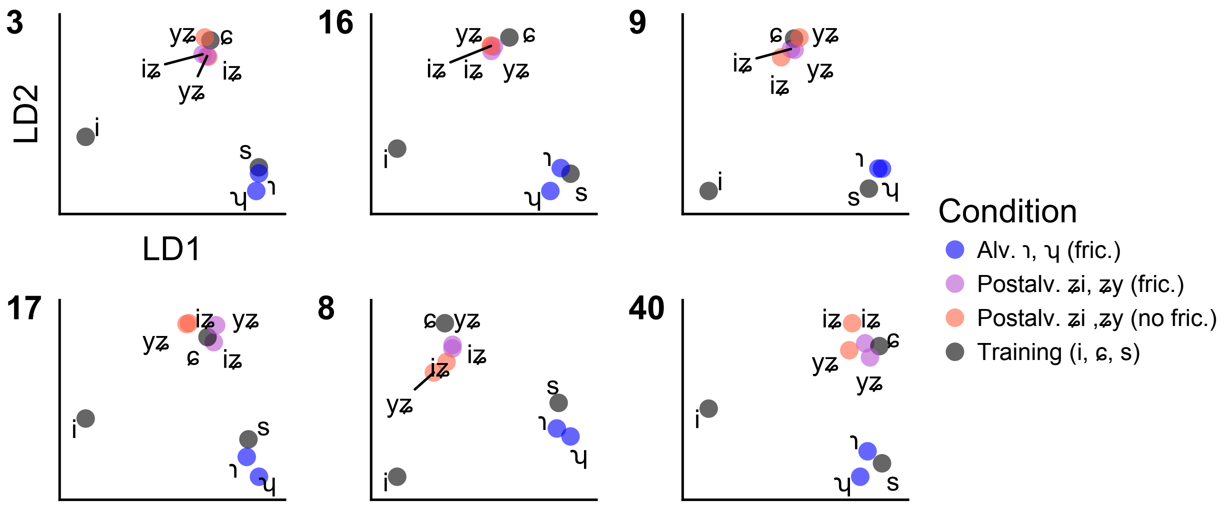


Figure 4.11: Median LD1-LD2 values for selected speakers by segment type and segmental context, illustrating noticeable articulatory uniformity among postalveolar segments.

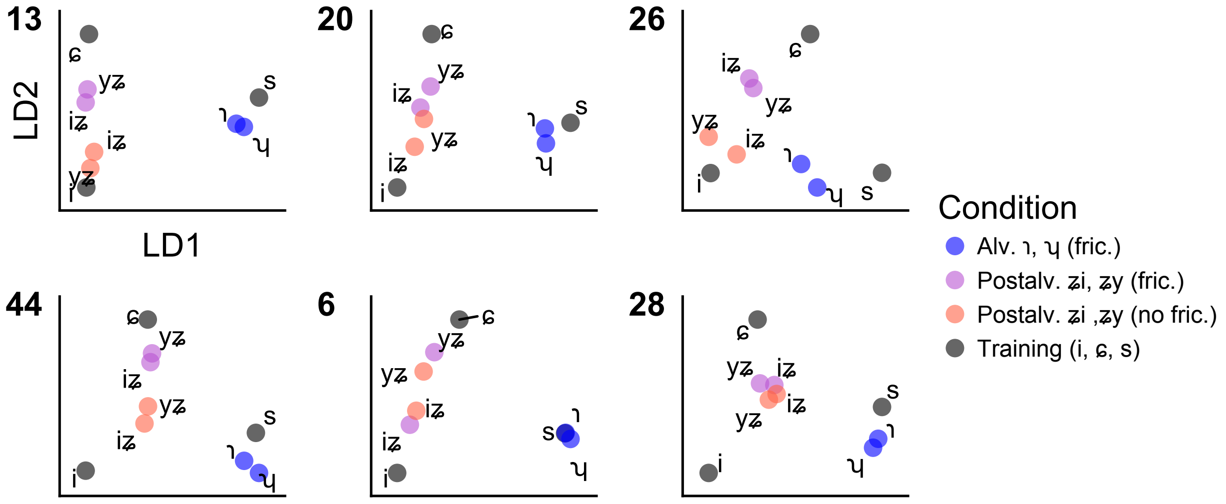


Figure 4.12: Median LD1-LD2 values for selected speakers by segment type and segmental context, illustrating a lack of articulatory uniformity among postalveolar segments; an effect of the presence of fricative onsets is often evident.

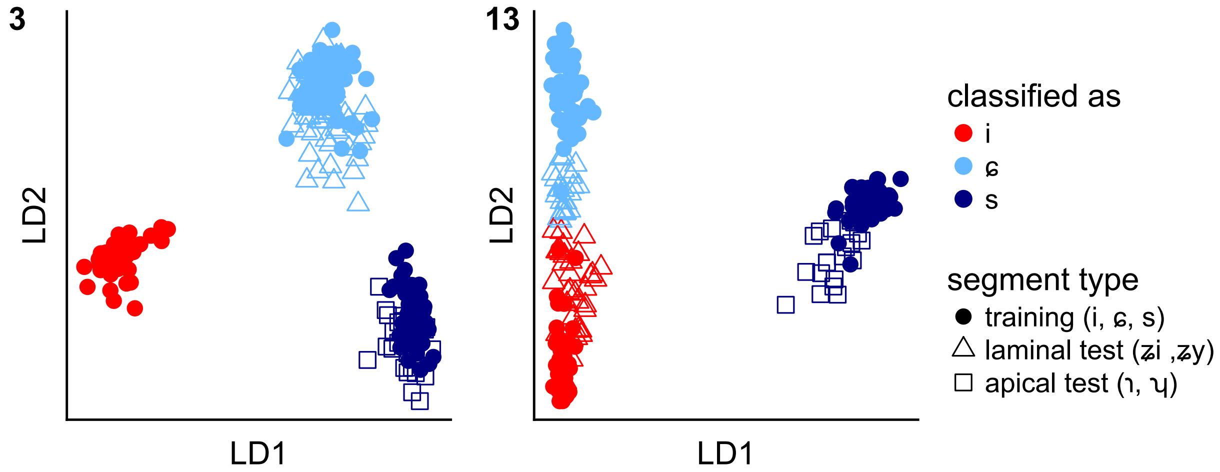


Figure 4.13: Example distribution in LD1-LD2 space and LDA-provided classification of all test items for two participants.

4.4.3 Articulatory similarity of fricative consonants and fricative vowels

From the analysis in the previous section, we know that the lamino-postalveolar vowels /i͡ʑ/, /y͡ʑ/ range in their lingual postures from relatively /i/-like to relatively /ɕ/-like or can even be indistinguishable from the lingual posture for /ɕ/. Having determined that there is some inter-subject variation in how much of the /i͡ʑ/ cluster is classified as /ɕ/, however, it remains to be determined how specifically similar, in a gradient way, /i͡ʑ/ is to /i/ on the one hand or /ɕ/ on the other hand. Given my impressions of potentially innovative variants used by younger speakers, it would not be surprising if the /i͡ʑ/ vowel became more /ɕ/-like in apparent time. If pressure towards motor program uniformity operates as a constant, low-level bias on articulatory outcomes, as generations of language acquisition pass in Sūzhōu Chinese, one would expect that population-level variability in the motor programs implemented for /i͡ʑ/ would gradually be winnowed down to a single typical articulatory strategy, in this case one resembling /ɕ/.

To evaluate similarity of the lamino-postalveolar vowels' lingual articulations to /i/ and /ɕ/, most similar to the conservative and innovative /i͡ʑ/, /y͡ʑ/ variants, respectively, ultrasound frames corresponding to the midpoints of all /i/, /ɕ/, and /i͡ʑ/, /y͡ʑ/ tokens were submitted to a separate PCA for each speaker, following the same procedure as described in the previous section but excluding the apico-alveolar vowels and /s/. These new principal components were used to train a new linear discriminant analysis on the training segments /i/ and /ɕ/; PCA scores for /i͡ʑ/ and /y͡ʑ/ tokens were then transformed to the resulting one-dimensional linear discriminant function. At one endpoint of the range of linear discriminant scores are those tokens of /i͡ʑ/ that are most like /ɕ/ and at the other endpoint are those instances of /i͡ʑ/ that are most like the /i/ training examples. The classification

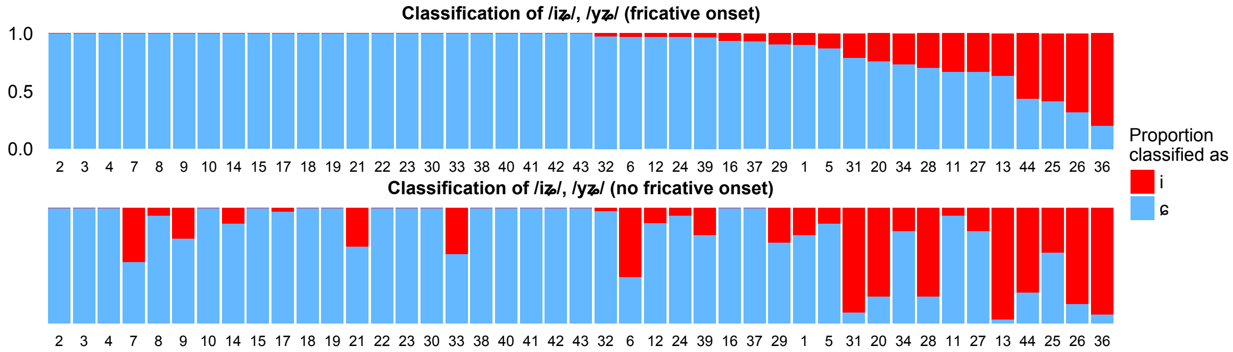


Figure 4.14: LDA classification of lamino-postalveolar test items only. Subject results are sorted by proportion classified as /ɕ/ in the fricative onset condition (top).

results of this LDA are displayed in Figure 4.14. These results are qualitatively similar to the results shown in Figure 4.10 in that most tokens of /i͡ʑ/ and /y͡ʑ/ are classified as /ɕ/, with some idiosyncratic variation from speaker to speaker.

Scores on this linear discriminant for the /i͡ʑ/ tokens were range-normalized within-talker to make them comparable across talkers. Using these normalized LD values, a ɕ-score was calculated by calculating the distance in this normalized LD1-LD2 space between each /i͡ʑ/ token and the median /ɕ/ value for the linear discriminant. The ɕ-score for a given token is calculated as:

$$\text{ɕ-}score = \frac{LD_T - \tilde{\mu}_{LD\text{ɕ}}}{\tilde{\mu}_{LD\text{i}} - \tilde{\mu}_{LD\text{ɕ}}}$$

where LD_T is the range-normalized linear discriminant value of the token, $\tilde{\mu}_{LD\text{ɕ}}$ is the median range-normalized linear discriminant value for that speaker's /ɕ/, and $\tilde{\mu}_{LD\text{i}}$ is the median range-normalized linear discriminant value for that speaker's /i/.

The ɕ-score provides a metric of distance (on the range-normalized linear discriminant) from /ɕ/ in the direction of /i/. If a token has a ɕ-score of 0, this indicates a production identical on the linear discriminant to the median LD for /ɕ/. The more similar to /i/ and less similar to /ɕ/, the more positive the ɕ-score, with a ɕ-score of 1 indicating a production identical to /i/ on the linear discriminant. A negative ɕ-score is possible and is attested for several speakers; this is the result of the LD scores for that speaker's /ɕ/ exhibiting a particularly wide distribution, such that the maximum LD score (one of the bases for range-normalizing the LD) is some distance away from the median (one of the bases for calculating the ɕ-score). Put another way, a negative ɕ-score suggests that the linear discriminant calculated to maximally separate /i/ and /ɕ/ happens to separate /i͡ʑ/ and /y͡ʑ/ from /i/ even more effectively. Median ɕ-score values for all subjects across all tokens in each onset context are shown in Figure 4.15.

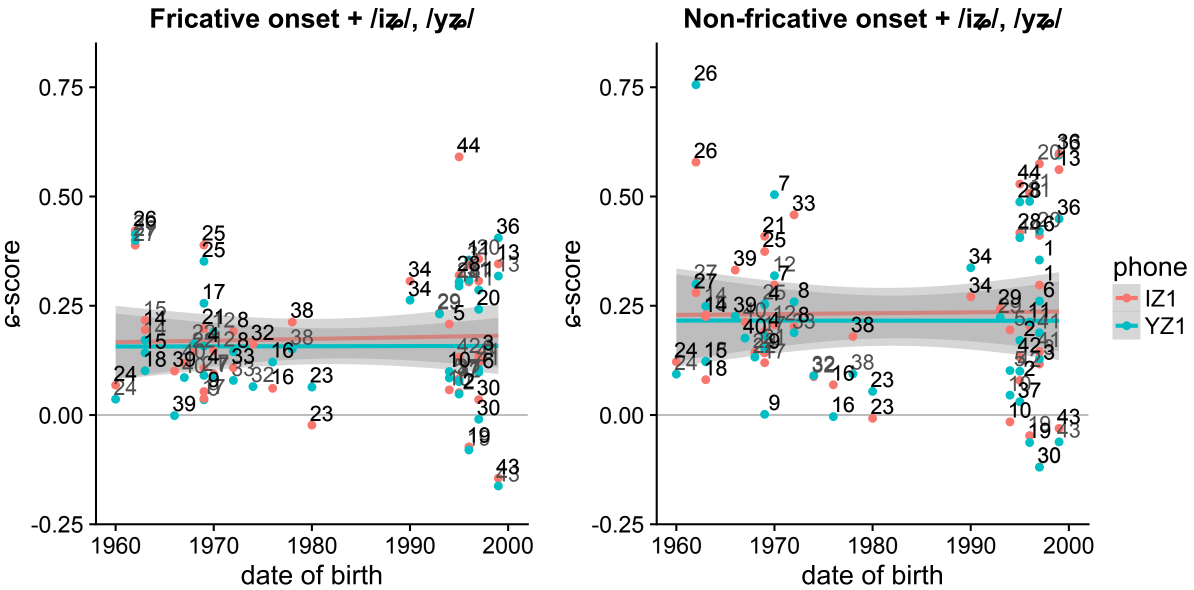


Figure 4.15: Median ɕ-score for each participant with respect to birth year, plotted with simple linear regression lines to visualize effects within the data. Data is presented in two groups, according to presence (L) or absence (R) of an onset fricative. Linear regression for each subgrouping by phone (/i͡ʑ/ or /y͡ʑ/) is overlaid.

To isolate the effect of age on overall ɕ-score, ɕ-scores for all /i͡ʑ/ tokens were submitted to a linear mixed effects model with fixed and interaction effects for age, phone (/i͡ʑ/ vs. /y͡ʑ/), and presence or absence of fricative onset, with random intercepts for subject (i.e., ɕ-score ~ age*phone*fric_onset + (1|subj)). The choice of factors for this model is hypothesis-driven, primarily by Hypothesis 2 (see Section 4.1.2), concerning whether similarity of /i͡ʑ/ and /y͡ʑ/ and to /ɕ/ increases over time. The exploration of LD-based spaces in Figures 4.8–4.9 suggests that fricatives in the segmental context and whether the vowel is rounded or not must be taken into account, given that they introduce a great deal of structured variation into each speaker's data set. The model was fit using the LME4 package (Bates et al., 2018) in R 3.5.0 (R Core Team, 2018). *p*-values for regression coefficients were calculated using the CAR package (Fox et al., 2018).

Model results are provided in Table 4.4. The main effect of age on ɕ-score fails to reach significance, suggesting that speaker age does not influence ɕ-score (see also Figure 4.15), counter to Hypothesis 2. Main effects of phone and fricative onset are significant, however. Presence or absence of a fricative onset is highly predictive ($p < 0.001$) of ɕ-score, as might be expected: the higher ɕ-score indicates less articulatory similarity to /ɕ/ when a fricative does not immediately precede the phone. The effect size is quite small, however, as can be confirmed by inspecting Figure 4.15. This suggests that similarity of /i͡ʑ/ and /y͡ʑ/ to /ɕ/ is relatively invariant and cannot be attributed to coarticulation with /ɕ/-like segments, consistent with Hypothesis 1b. The choice of /i͡ʑ/ and /y͡ʑ/ as the phone is also highly sh_score ~ age*phone*fric_onset + (1|subject)

Random effects	Variance	St. dev.			
subject	0.02271	0.1507			
Residual	0.01644	0.1282			
Fixed effects	Estimate	St. err.	*t*	p $(> \chi^2)$	
(Intercept)	0.2159714	0.0625673	3.452	—	
age	-0.0008727	0.0016545	-0.527	0.93869	
phone	-0.063388	0.0201941	-3.139	6.59e-06	***
fric_onset	0.003771	0.016899	0.223	2.20e-16	***
age:phone	0.0010103	0.0005349	1.889	0.31124	
age:fric_onset	0.0011994	0.0004445	2.699	0.03319	*
phone:fric_onset	0.0521167	0.028775	1.811	0.46717	
age:phone:fric_onset	-0.0012588	0.0007584	-1.66	0.09694	.

Table 4.4: Results of linear mixed-effects regression across all tokens regardless of onset fricative condition (present or absent).

predictive: /y͡ʑ/ yields a slightly lower ɕ-score than /i͡ʑ/, suggesting that of the two vowels, /y͡ʑ/ is on the whole slightly more similar in articulation to /ɕ/.

In the model shown in Table 4.4, an interaction effect of age and fricative onset also reaches significance, and the three-way interaction of age, fricative onset, and phone approaches (but does not reach) significance. To better understand the nature of these interactions, a follow-up analysis was carried out. The data used for the first analysis was subsetted into two groups, the data with fricative onsets and the data without fricative onsets, corresponding to the two levels of the fric_onset factor. Separate linear mixed-effects regressions were performed on each subset of the data for age, phone, and their interaction, with random intercepts for subject; models were fit and their effects' significance levels evaluated using the same method described above.

The resulting pair of models appears in Table 4.5. A significant main effect of phone is present for both conditions, although it is more predictive for the fricative onset condition ($p < 0.001$) than the non-fricative onset condition ($p < 0.05$). For the portion of the data with fricative onsets, an interaction effect of age and phone also reaches significance. This suggests that in a fricative-onset context, /y͡ʑ/ is slightly more /ɕ/-like for younger speakers (and slightly less /ɕ/-like for older speakers); this interaction may also drive the nearly significant three-way interaction observed in Table 4.4. The models separated by fricative onset condition do not, however, offer an explanation for the weakly significant interaction between age and fricative onset condition in the main model (Table 4.4): the main effect of

Onset fricative; sh_score ~ age*phone + (1|subject)

Random effects	Variance	Std. dev.			
subj	0.018663	0.13661			
Residual	0.009654	0.09826			
Fixed effects	Est.	Std. err.	t	p ($> \chi^2$)	
(Intercept)	0.2101053	0.0564702	3.721	—	
age	-0.0007965	0.001493	-0.533	0.75878	
phone	-0.0572404	0.0156936	-3.647	1.73E-05	***
age:phone	0.0009166	0.0004138	2.215	0.02676	*

No onset fricative; sh_score ~ age*phone + (1|subject)

Random effects	Variance	Std. dev.			
subj	0.03142	0.1772			
Residual	0.01794	0.1339			
Fixed effects	Est.	Std. err.	t	p ($> \chi^2$)	
(Intercept)	0.2208973	0.0731938	3.018	—	
age	0.0003051	0.0019363	0.158	0.91029	
phone	-0.0099215	0.0214573	-0.462	0.01472	*
age:phone	-0.0002666	0.0005624	-0.474	0.63549	

Table 4.5: Results of linear mixed-effects regression repeated across both onset fricative conditions, present (top) and absent (bottom).

age in each model in Table 4.5 fails to reach significance.

4.5 Discussion

On the whole, the results of the production experiment undertaken in this chapter have shown that speakers of Sūzhōu Chinese exhibit articulatory strategies for the postalveolar fricative vowels /i͡ʑ, y͡ʑ/ that mainly resemble /ɕ/. Speakers also use production strategies for the apico-alveolar fricative vowels /ɿ/, [ʮ] that resemble /s/. This general uniformity of articulatory strategy is revealed both by data-driven classification and exploratory data visualization of reduced-dimensionality representations of collected ultrasound data. Individual speakers are also seen to deviate from the main tendency to have uniform productions across the fricative vowels and fricative consonants produced at a given constriction location, consistent with use of the tongue dorsum as the active articulator (producing a dorso-postalveolar vowel) instead of the tongue tip or blade.

Articulatory similarity of the alveolar fricative vowels /ɿ/, [ʮ] to alveolar fricative or affricate segments such as /s/ is expected, owing to co-occurrence restrictions that require /ɿ/, [ʮ] to occur immediately following an alveolar fricative or affricate. The Sūzhōu Chinese postalveolar fricative vowels /i͡ʑ̩, y͡ʑ̩/ are not subject to such a restriction, occurring both immediately after fricatives and elsewhere, allowing for uniformity to be separated from coarticulatory pressures as a source of similarity in articulatory strategy. Modeling the ultrasound data reveals that presence or absence of a fricative onset is predictive of changes in articulatory strategy for /i͡ʑ̩/ and /y͡ʑ̩/, but that the effect on a metric of /ɕ/-likeness (the ɕ-score) is modest. Demographic variables such as age and gender also do not appear to predict trends in ɕ-score, suggesting that individual deviations from the main pattern of uniformity are idiosyncratic.

Below, some implications of the findings are briefly discussed (Sections 4.5.1–4.5.2), particularly for the existing body of research on uniformity (Keating, 2003; Ménard et al., 2008) and structured variation (Chodroff, 2017; Chodroff and Wilson, 2017), but also for theories of sound change and their propagation across populations (Ohala, 1989; Ohala, 1993). To conclude, some shortcomings of the present research and the possibilities of future research on Wú Chinese are offered in Section 4.5.3.

4.5.1 Uniformity and the "pool of synchronic variation"

The analyses in Sections 4.4.2–4.4.3 suggest, counter to initial expectations, that a variety of articulatory strategies—uniform and non-uniform alike, though uniformity predominates—for producing /i͡ʑ̩/ and /y͡ʑ̩/ are present in all generations of the Sūzhōu Chinese-speaking population examined. These strategies range from a /ɕ/-like lingual posture that is relatively uniform with the set of alveolopalatal fricative-producing sounds to /i/-like variants that have motor programs intermediate between /ɕ/ and /i/ (and uniform with neither). There does not appear to be a trend toward more prevalent uniformity in apparent time, and some less /ɕ/-like variants are actually more in use with younger speakers, according to the analyses in Section 4.4.3.

However, some degree of speaker idiosyncrasy in coarticulation of fricative vowels and onset fricative consonants is evident in the data. To underscore the type and degree of idiosyncrasy present in the data, median ɕ-score data are presented again in a modified format in Figure 4.16. Many speakers can be seen to switch between relatively /ɕ/-like and /i/-like strategies for the postalveolar fricative vowels depending on the context: relatively /ɕ/-like when following a fricative onset consonant, and relatively /i/-like elsewhere. This could be attributed to automatic coarticulation with the preceding segment. For some speakers, however, the opposite appears to occur, with their more /i/-like productions occurring in the fricative onset context. The degree to which the articulations of /i͡ʑ̩/ and /y͡ʑ̩/ respond to coarticulation also varies substantially from subject to subject.

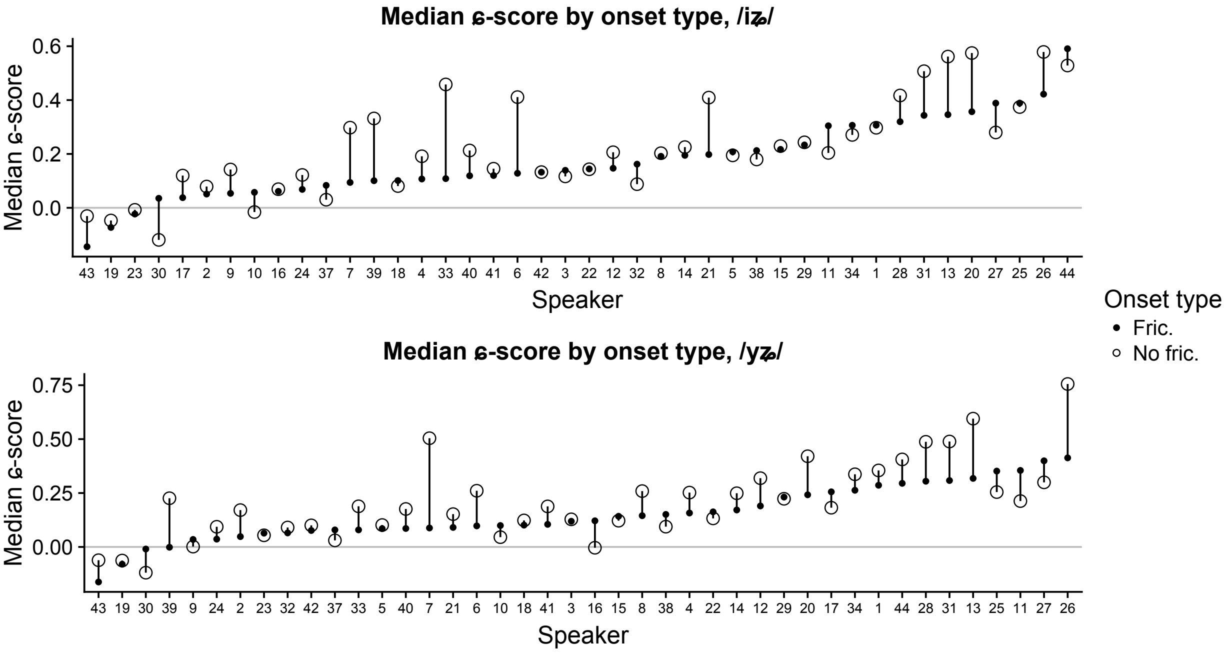


Figure 4.16: Median ɕ-scores by condition and vowel, joined by vertical lines, for each subject. Each vowel's data is sorted by increasing median ɕ-score in the fricative onset context. Note that in the presence of an onset fricative, some subjects decrease their typical ɕ-score, some exhibit little to no change, and several increase their typical ɕ-score.

Nonetheless, most speakers have a relatively low ɕ-score for their postalveolar vowels /i͡ʑ/, /y͡ʑ/, suggesting that uniform articulations predominate. It stands to reason that uniformity should contribute to—and give a distinctive bias to—the Ohalian "pool of synchronic variation" (Ohala, 1989) from which sound changes are drawn. However, no population-level sound change appears to be in progress in spite of the most frequent tendency overall being uniform production: as modeled in Tables 4.4–4.5, speakers of all ages are similarly likely to choose non-uniform production strategies. The "pool of variation" is thus relatively stable across time periods. Why do uniform production strategies not make up a larger share of the "pool" with the passage of several generations of language acquisition? In Sūzhōu Chinese, the consequences of selecting the uniform strategy across fricative consonants and fricative vowels appear to be limited to the speaker: they may not be perceived by other Sūzhōu Chinese speakers.

Some further elaboration is required on this point, since formant frequencies, particularly F2, clearly vary among speakers (see Figure 4.3), and are possibly systematically related to fricative vowel articulatory strategy. I speculate that while formant frequency differences among the postalveolar fricative vowel variants are *perceptible*, Sūzhōu Chinese speakers do not *attend to* them. Interspeaker differences in typical formant frequencies, particularly F2, are measurable in production of /y͡ʑ/ and (especially) /i͡ʑ/. For the task of comprehending

Sūzhōu Chinese as a native speaker, however, formant frequencies may not be especially reliable as cues to these vowel categories, given crowding of the high vowel space and the fact that vowels with places of articulation more anterior than [i]'s dorso-palatal constriction have different acoustics-articulatory relations than dorso-palatal vowels such as [i], [ɨ], or [u] (see Section 3.4.1), but occupy the same range of F1-F2 space that they do. Fricative noise may thus be a more reliable cue for these non-dorso-palatal vowel categories.

Furthermore, while the choice between a dorso-postalveolar or lamino-postalveolar vowel may have some consequences for formant frequencies, the spectral profile of the fricative noise that results seems to be broadly similar between the two strategies, given the similar constriction location that results (Ling, 2009). Further research on the perceptual organization of the Sūzhōu Chinese vowel space may clarify the relative importance of fricative noise and formant frequencies in cueing the fricative vowels, and whether Sūzhōu Chinese speakers attend to variation in F2 in the postalveolar fricative vowels more specifically.

The Sūzhōu Chinese-specific pattern of uniformity discussed here is thus likely a sound pattern that cannot "koinéize" (Mielke et al., 2016), spreading to a user base larger than the individual speaker and losing its individual-level complexities in the process. Individual grammars, which are variously more or less influenced by uniformity as a desideratum, cannot converge on a community norm if (as in this case) the consequences of uniformity are not perceptible. However, owing to the tendency to reuse articulatory primitives during language learning, uniform articulatory strategies are presumably ubiquitous in speaker repertoires. There is also no reason to assume that articulatory uniformity's influence on the implementation of phonological primitives can never be perceptible. Articulatory uniformity can thus be taken to constitute a distinctively "speaker-driven" element of synchronic variation that may contribute to advancing sound changes in a community. This presents a complication to theories of sound change that argue that all sound change is listener-driven (Ohala, 1989; Ohala, 1993). More discussion on uniformity and its potential role in driving sound change, can be found in Chapter 5.

4.5.2 Covert articulatory uniformity

An important aspect of the uniformity between fricative vowels and fricative consonants in Sūzhōu Chinese is that it is largely covert: the choice between a relatively dorso-postalveolar and a relatively lamino-postalveolar articulation strategy for /i͡ʑ/ and /y͡ʑ/ does not appear to have a large, consistently perceptible effect on the quality of the frication produced, as argued in the previous section. This is in contrast to many existing case studies of uniformity, which most often focus on uniformity of acoustic outputs. In particular, recent research on "structured variation" focuses on choice among articulatory strategies with overt, distinct acoustic consequences. For instance, covariation of stop VOT across stop place has recently been used as a gauge of uniformity of timing in laryngeal articulation (Chodroff and Wilson, 2017; Chodroff, 2017). Chodroff (2017) also evaluates uniformity of voiced and voiceless fricatives in terms of a single acoustic dimension: mid-frequency spectral peak, or Freq_M (Koenig et al., 2013; Shadle et al., 2016). Each of these studies rests on the reasonable assumption that the acoustic parameters measured relate to some aspect of articulation: VOT relates to timing patterns in laryngeal articulation, and Freq_M relates to the length of the front cavity of a fricative.

Other documented cases of articulatory uniformity, however, are ill-reflected in acoustic outputs, which may themselves be non-uniform. Some attributes of laryngeal articulation that determine the timing of positive and negative VOT covertly vary (Flege, 1982; Keating, 2003); speakers occasionally exhibit articulatory uniformity in a covert way by consistently employing articulatory strategies that do not have a substantial impact on the acoustic output. Likewise, Freq_M is a reliable index of constriction anteriority, but there are multiple ways of achieving the same constriction location, as is in fact demonstrated by the Sūzhōu Chinese data here. The postalveolar fricative vowels may have either the tongue dorsum or the tongue blade or tip as the active articulator, and the resulting frication appears to have comparable spectral attributes in either case.

As discussed above, the model of uniformity-through-reuse would seem to predict that uniform articulatory strategies for series of segments should come to dominate at the population level over time, given increasingly uniform input to language acquisition from the speakers in the learner's environment. Uniform strategies, given that they approximately fulfill the needs of the speaker for a given task and are more easily mastered, would seem to be akin to a "dominant gene" that should eliminate non-uniform strategies in the speaker population over time. The lack of change over time in the rate at which speakers use uniform strategies would seem, at first glance, to be a strike against another prediction of uniformity-through-reuse.

In perceiving the fricative vowels, speakers may attend to fricative noise in the spectrum, which may be a target for fricative vowel production and does not differ substantially between uniform and non-uniform strategies. Any difference in *formant frequencies* between uniform and non-uniform strategies may not be perceptually important. A uniform articulatory strategy in the Sūzhōu Chinese fricative vowels may thus not be inherited like a "dominant gene," instead arising through speaker-side factors alone. Acquisition in these cases is akin to rolling the dice, which will tend to result in uniformity but occasionally not, due to idiosyncrasy; thus, some low degree of variation in articulatory strategy persists across the population.

The data collected here on Sūzhōu Chinese is thus a contribution to the growing portion of the field of phonetics and phonology that recognizes that not all sound patterns that are of interest to theoretical linguistics are based on perceptible acoustic outputs (Mielke et al., 2016). Further study of articulatory uniformity stands to provide valuable information on the details of phonetic implementation of phonological primitives, and perhaps even clarify the nature of phonological primitives themselves.

4.5.3 Caveats and prospects: complexity of the Sūzhōu Chinese sociolinguistic situation

The complexity of the Sūzhōu Chinese sociolinguistic situation may mask patterns in the articulatory strategies used for /i͡ʑ/, /y͡ʑ/ over time. As discussed in Section 4.2.1, language history and residential history are largely homogenous for the study participants taken as a whole. However, particularly for participants under the age of 30, the study population can be said to contain several lurking variables that are more difficult to quantify than the major biographical characteristics collected in the present study. The actual extent to which speakers use Sūzhōu Chinese and Standard Chinese in their daily lives is difficult to gauge, as is their orientation toward the local dialect and tradition, as opposed to Standard Chinese, modernity, and cosmopolitanism. More fundamentally, an objective way of rating proficiency in Sūzhōu Chinese is needed. Speakers were asked in a metadata collection survey to provide self-ratings of their ability in Sūzhōu Chinese and Standard Chinese. However, these ratings appear to be unreliable, and the self-rating task subject to misinterpretation;[9] as such, the self-ratings were not used here.

If these lurking variables can be better accounted for, more predictive power may be attained with respect to changes (or the lack thereof) in articulatory strategy for the Sūzhōu Chinese postalveolar vowels. Younger speakers, as a group, typically use a wider variety of articulatory strategies compared to older speakers (see Figure 4.15), which may relate to the lurking variables described above. In particular, the level of fluency may vary among this group more substantially than for older speakers due to a higher rate of use of Standard Chinese, which may be reflected in the wider range of ɕ-score measurements for this group. Several participants around the age of 20 had difficulty with consistently producing an expected Sūzhōu Chinese reading for the stimulus hànzì (e.g. Speakers 19, 36). Speaker 35, one of the youngest speakers in the study, merits special note for lacking a systematic contrast between /i͡ʑ/ and /i/ altogether. Young Shànghǎi Chinese speakers exhibit Speaker 35's pattern as a rule (Zhu 2006); contact with Standard Chinese is considerably more intense in Shànghǎi than in Sūzhōu and has presumably contributed to the merger of this unusual contrast.

Future research on speech production in Wú Chinese must take this complex social situation fully into account. On the one hand, this means that extracting generalizations from data on Wú Chinese varieties is bound to be rather difficult, owing to the heterogeneity of the speaker population, its social circumstances, and its attitudes towards the linguistic situation shared in common by non-standard Chinese speakers. On the other hand, this underscores that Wú Chinese varieties, which are strikingly understudied relative to the size and accessibility of the speaker population and contain a range of unusual phonetic phenomena, are potentially a source of empirically rich data on the interactions between paradigmatic

---

[9] One participant, for example, communicated with the experimenter in fluent Standard Chinese and experienced noticeable difficulty with producing consistent Sūzhōu Chinese readings in the production task. Nonetheless, on the survey's 7-point Likert scale, in which "1" indicates poor speaking ability and "7" fluent speaking ability, he rated himself at "1" for Standard Chinese but "6" for Sūzhōu Chinese.

phonological factors, social factors, and phonetic implementation. The development of the postalveolar fricative vowels across Wú Chinese varieties merits further study in particular.

Chapter 5

Discussion and conclusions

In this concluding chapter, I begin by summarizing the theoretical framework and findings of the dissertation (Section 5.1). I then discuss implications of this dissertation's research and provide directions for future studies (Sections 5.2–5.3).

5.1 Brief review of the dissertation

This thesis has provided a comprehensive review of uniformity and similar principles in linguistic science; to my knowledge, this is the only review of such topics that incorporates evidence from both linguistics (phonetics, phonology, and their interface) and motor control (speech motor control and otherwise). This review has aimed to provide an explanation for the observations that sounds are organized into series that share phonological primitives among themselves, and that these sounds exhibit constrained variability in the phonetic dimensions that relate to these primitives. Below, as a prelude to discussing future directions starting in Section 5.2, I review the major aspects of the model of uniformity-from-reuse developed in this thesis, and how the results of the experiment in Chapter 4 both inform and complicate this model.

5.1.1 Uniformity and its origins in articulatory reuse

I describe UNIFORMITY as the pattern by which phonological primitives map to maximally similar outputs when implemented across a series. This framework assumes as basic a division between mental representation of action (generated by a "phonology" module, as discussed here) and action itself (a "phonetics" module) (Keating, 1984; Keating, 1990). The apparent behavioral root cause of uniformity, ARTICULATORY REUSE, occurs during language development. Reuse of task-directed actions, or components of actions, is known to result in success at a related goal—here, related in perceptual terms—and is an efficient means of learning new speech motor tasks in the absence of a detailed, reliable INTERNAL MODEL (Wolpert et al., 1995; Wolpert and Flanagan, 2001; McAllister Byun et al., 2016) of how articulation relates to acoustic results. In Chapter 2, I have elaborated a model of articulatory reuse and articulatory uniformity and provided justification for connecting these phenomena: reusing articulatory strategies may result in functional connections among speech sounds that may persist into adulthood.

The case study of uniformity in Sūzhōu Chinese, while sufficiently involved to require some detailed description (Chapter 3), is advantageous in two ways. First, and more obviously, the data concerns the phonetic structures of a relatively underdocumented language; these phonetic structures themselves are underdocumented, particularly the FRICATIVE VOWELS. Secondly, and less obviously, the fricative vowels themselves occupy a structural position in Sūzhōu Chinese and other languages with fricative vowels (see Chapter 3) which may offer a window into the factors that mediate uniformity. Speakers of Sūzhōu Chinese favor articulating the fricative vowels in one of two ways, with dorsal constrictions (somewhat akin to the high front vowels) or uniformly with consonant sounds made with non-dorsal constrictions.

5.1.2 Experimental findings from Sūzhōu Chinese

The aim of the Sūzhōu Chinese production experiment described in Chapter 4 was to confirm whether uniformity holds among Sūzhōu Chinese's fricative consonants (e.g., /ɕ/) and its fricative vowels (/i͡ʑ̩/, /y͡ʑ̩/). Through direct assessment of ultrasound data, most speakers were found to favor a uniform articulatory strategy across the fricative consonants and vowels speakers. This is in keeping wih the model outlined above, in which speech sounds that present similar acoustic targets to the learner may be learned as sharing a single articulatory primitive, regardless of the articulatory nature of the sounds in the model. In the Sūzhōu Chinese case specifically, uniformity could arise from reuse of the tongue position from the /ɕ/ series of consonants for production of a common acoustic goal ([ɕ]-like fricative noise) in fricative vowels.

However, some speakers favor a non-uniform production strategy, with the fricative vowels taking on a lingual posture intermediate between the /ɕ/ series and a high front vowel such as /i/. This recalls other case studies of articulatory uniformity, where not all speakers examined are uniform in their articulation: Keating (2003), for instance, makes it clear that uniformity is violable, but nonetheless prioritized by most of the study population in Flege (1982). Of more interest is the fact that uniform strategies for fricative vowel production do not come to predominate in the population over time. This may occur because selecting the uniform strategy for fricative vowels may not result in a change perceptible to the speaker of Sūzhōu Chinese; an imperceptible variant cannot generalize to the population to become a more general pattern (Mielke et al., 2016). However, is it not necessarily the case that all uniformity must face this restriction simply because it occurs in Sūzhōu Chinese; uniformity's possible role in directing and constraining sound change is discussed below.

5.2 Articulatory uniformity and sound change

Articulatory reuse has implications for the development of some phonological structures in that it potentially removes the specifically linguistic impetus for such developments. That is, the research carried out in this dissertation suggests that certain learning processes—which are not specifically linguistic, but are a feature of learning in biological organisms more generally—may cause articulatory uniformity at the individual level and the development of phonological series at the community level. Section 5.2.1 is given over to this idea, which would displace purely linguistic motivations such as Feature Economy (Clements, 2003). There are also implications for typologies of sound change and the source of population-level biases in the "pool of synchronic variation" (Ohala, 1989), since uniformity may exert substantial influence on the structure of this "pool" (Section 5.2.2).

5.2.1 From uniformity to higher-level phonological structures

Writing on Feature Economy, Clements (2003) observes:

> Phonology can be seen as a grammaticalisation of the quantitative patterns determined by phonetic constraints, and feature economy can be viewed, in some instances, as a grammaticalisation of gesture-economy effects operating at the more abstract feature level. Thus, non-distinctive allophonic patterns introduced in particular contexts may generalise over classes of sounds due to gesture economy, and if these patterns subsequently become grammatically or lexically relevant ... a pattern that was originally determined by gesture economy will have become transformed into a pattern of feature economy. (Clements, 2003, pp. 326–27)

Put another way, a driving force behind feature economy may be the tendency toward gestural economy, since it contributes to the formation of groups of segments similar enough for learners to induce natural classes from them. Gestural economy is readily identifiable with articulatory uniformity as described in Chapter 2; I speculate here that articulatory uniformity and reuse may play some role in the development of series and other high-level phonological structures.

Segments in series constitute part of the typical state of affairs in an average language's phonological inventory. It has been remarked that series are typical and so-called ISOLATED SOUNDS are relatively uncommon in the languages of the world (Martinet, 1955; Clements, 2003). For purposes of the present discussion, isolated categories can be defined as those categories that do not share some number of phonological features with other phonemes in their language, or which do not neatly fit into a series. The rarity of isolated phonological categories relative to non-isolated categories is a frequent source of discussion in early structuralist accounts of economy, perhaps most saliently Martinet (1955). Martinet attributes relatively infrequent attestation of isolated categories to a certain precariousness of

Stage 1				Stage 2			
ejective	p’	t’	k’	ejective	p’	t’	k’
voiceless aspirated	p^h	t^h	k^h	voiceless aspirated	p^h	t^h	k^h
voiced	b	d	g	*voiceless*	p	t	k
voiceless			k	*voiced*	b		g
implosive	ɓ						

Table 5.1: The Zulu oral stop inventory over time, from Clements (2003). Stage 1: the inventory recorded by Doke (1963) contains isolated sounds (shaded cells). Stage 2: more recent studies (Traill et al., 1987; Best et al., 2001), the two formerly isolated stops form a voiced series contrastive with the formerly voiced series, which is now voiceless.

their position within the phonological system, and explicitly invokes the outcomes frequently attested in sound change:

> All of this confirms […] the instability of isolated, non-integrated phonemes in a system. They are, in effect, more prone to disappearing, to creating a correlative partner, or to evolving to act as a partner of another isolated phoneme. (Martinet, 1955, p. 80 [1])

Languages do not tend to avoid isolated segments, in other words, so much as they tend to lose them, and the evidence reviewed in this dissertation suggests articulatory reuse as one reason for this loss. Isolated sounds are less likely to be accurately transmitted as isolated sounds from one generation of speakers to another, and more likely to “fall into” a pre-existing series for a given generation of speakers, because articulatory reuse persistently pushes learners in this direction.

For instance, Clements (2003) reviews evidence that sound change of the Zulu stop consonants has created a contrastively voiced short-lag stop series (Table 5.1). Evidence for two formerly isolated sounds, an implosive stop /ɓ/ and a voiceless unaspirated stop /k/, can be found in early documentation of Zulu (Doke, 1963); these have each changed in their typical productions and given rise to a new voiced series (see Traill et al., 1987; Best et al., 2001). Numerous other examples of this sort of series-producing or enhancing sound change can be found in Martinet (1955) (see pp. 59–70).

Articulatory reuse, which gives rise to uniformity in individual speakers, could ultimately also be the root cause for large-scale phonological structures of the sort attributed to Feature Economy in the examples above. Developing an articulatory strategy through reuse facilitates learning (Loeb, 2012); speakers may bootstrap off of the resulting increased similarity

---

[1] Translated from the original: *Tout ceci confirme […] l’instabilité des phonèmes isolés, non intégrés, dans un système. Ils sont, en effet, plus exposés à disparaître, à se créer un partenaire corrélatif ou à évoluer pour servir de partenaire à un autre isolé.*

and induce the existence of phonological classes on similarity grounds (Mielke, 2008). Iterated over many generations of language acquisition (this being a kind of learning scenario where non-optimal learning may prevail), biases towards producing similarity (and then finding and perhaps enhancing it) could lead to large-scale phonological structures. Notably, and as is not the case for Feature Economy, nothing in this framework is specifically linguistic: large-scale phonological structures could be viewed as merely a side effect of the way biological organisms engage in task learning, and not a linguistic desideratum specifically.

5.2.2 Uniformity as "speaker-driven" sound change

Research on articulatory reuse and uniformity presents an interesting complication to typologies of sound change. The dominant Ohalian model holds that sound change is mainly driven by perceptual errors on the part of the listener (Ohala, 1993). The intended percept may fail to reach the listener either because the listener attributes variability in the signal to the intended signal itself (hypocorrection) or overcorrects the signal to arrive at a percept lacking some aspect of the intended signal (hypercorrection). Such LISTENER-DRIVEN sound change is generally adopted by more recently articulated theories of sound change as the major mechanism by which sound change occurs (Blevins, 2004; Garrett and Johnson, 2013).

Articulatory reuse (and its broader effects in the form of uniformity, described in this dissertation) could be viewed as a form of speaker-driven sound change, and more specifically a "mini" sound change (Ohala, 1993; Beddor, 2009) that initially affects only the individual. Rather than the root cause of this change being misperception, it is (in a sense) misproduction. As a result, the "pool of synchronic variation" frequently alluded to in Ohala's work (e.g. Ohala, 1989) has additional speaker-driven structure present; one could argue that a substantial amount of the bias within the pool of variation is structured by "speaker-side" factors like articulatory uniformity.

Beyond biases added to the pool of variation, uniformity is probably not without its impacts on sound structure, even if the uniform variant itself is only covertly different from some non-uniform variant(s). For instance, it has been observed that North American English speakers vary in the extent to which they favor or disfavor retroflex /ɹ/ variants in various segmental contexts (Stavness et al., 2012; Mielke et al., 2016). Given simple biomechanical constraints, such as the placement of the alveolar ridge between the typical tongue tip locations for interdental [θ] and retroflex [ɻ], these covertly retroflex variants may result in these speakers producing /θɹ/ sequences with an intervening tap (e.g., [θɾɻ]) at a greater rate than speakers who favor a bunched /ɹ/ variant. Thus, an otherwise covert variation in /ɹ/ type may have audible impacts on the articulation of other segments, possibly leading to the development of new phonological complexity (eventually, perhaps, allophony).

One can imagine a generalized case (Figure 5.1) in which articulatory reuse might bias speakers to produce covertly uniform variants of a speech sound A. These variants of A, or A', may make a secondary change in a neighboring segment B more likely, such that an audible consequence C may be detected. The result to the listener is that the sequence AB

Covert variation of A:	$A \sim A'$	[ɹ] ∼ [ɻ]
A does not condition $B \rightarrow C$:	AB, $*AC$	[θɹ], *[θɾɹ]
but A' conditions variation $B \sim C$:	$A'B \sim A'C$	[θɻ] ∼ [θɾɻ]

Figure 5.1: General example of covert variation driving overt variation (left) and a speculative example using North American English post-interdental /ɹ/-tapping (right).

loses ground to the apparent sequence AC (which is really covertly different $A'C$) in the community if uniformity is likely to make A' occur at some small rate. However, for the time being, this must be regarded as speculative. Future research may demonstrate that these sorts of effects can be attributed to covert variation generally, as in American English /ɹ/-tapping, or specifically to covert variation driven by articulatory uniformity, for instance if uniform variants of the Suzhou Chinese vowels differentially affected the realization of adjacent onsets (i.e., palatalization of onset /s/ to [ɕ], affrication of /t/ to [ts], etc.).

5.3 Extending the present research

The research here has focused on aftereffects of articulatory reuse during L1 learning in adult L1 phonologies, as in the Sūzhōu Chinese experiment in Chapter 4. Several types of studies not carried out here have the potential to help flesh out the theoretical framework of reuse-leading-to-uniformity. First, I propose to directly examine the articulatory strategies used by early language learners (Section 5.3.1). Secondly, I propose studies that examine whether L2 learners are, on occasion, also subject to the pressure to engage in articulatory reuse, as L1 learners are thought to be (Sections 5.3.2–5.3.3).

5.3.1 Reuse in L1 learning

If articulatory reuse is thought to predominate during language acquisition, then it stands to reason that collecting articulatory data at the right points in time during language acquisition may shed some light on whether (and if so, how) an individual child learner acquires uniform articulatory strategies. A longitudinal study of child learners and their articulation of a series of sounds—such as multiple vowel heights across different combinations of backness and rounding, as studied in Ménard et al. (2008)—could be used to develop a "before and after" snapshot of the child's progress. The "before" may show the child producing in non-uniform fashion, while the "after" may show the development of a uniform series based on the same articulatory primitives.

Actually observing pre-uniform articulatory strategy may be quite impractical, however, given that some uniform strategies are already cemented by age 4 (Ménard et al., 2008). The participants in a study of the sort proposed here may thus need to be even younger than this, presenting additional difficulties to experimenters in acquiring informative articulatory data in sufficiently large quantities to answer research questions. Other, more convenient means of observing articulatory reuse in action may be possible, which are explored below.

5.3.2 Reuse in L2 learning

Somewhat more speculatively, some of the same patterns of articulatory reuse that occur in L1 acquisition may occur in limited fashion in adult L2 acquisition. Unlike child learners, adults have no shortage of speech motor ability and experience, so many of the constraints discussed in Chapter 2, such as a generally unreliable internal model, do not apply to adult L2 learners. Furthermore, most L2 phones will also be readily identified with L1 phones that can pass straightforwardly for them, as predicted by the Perceptual Assimilation Model (PAM) of cross-language speech perception (Best et al., 1988; Best, 1994). Thus, adult L2 learners also have more material off of which to bootstrap in attempting to produce novel lexical content from another language.

However, depending on their L1, even adult learners will be unpracticed at applying certain types of controls to the speech production apparatus. For instance, front rounded vowels such as [ø] or [y] are difficult for English speakers to produce contrastively from back rounded vowels such as [o] or [u] (Flege and Hillenbrand, 1984; Levy and Law, 2010); this is presumably because they require a coordination of tongue body constriction and lip rounding that is not used to produce English vowels (Wood, 1986). It stands to reason that L2 learners might simply explore the possibilities of their L1 task space before developing new L2 articulatory strategies from scratch, reusing their L1 articulatory habit in the process. In this way, one can assume that for certain phones, adult acquisition shares some similarities with child acquisition.

L2 learners are generally observed to achieve some success in mimicking or drifting towards native-like L2 categories (Flege et al., 1997; Sancier and Fowler, 1997). However, an accumulation of experimental evidence suggests that adult learners never (Flege and Hillenbrand, 1984; Scovel, 1988) or rarely (Birdsong, 2007) achieve truly native-like performance in producing some or all L2 phonological categories. Scovel (1988) explicitly attributes this performance gap to neuromuscular habituation to L1 articulatory strategies. The direct articulatory evidence for claims of this sort is scant, however: persistent production differences in L2 learners tend to be reported in terms of perceptual judgments by native speakers of the language in question or acoustic measures. In part because these findings are illustrated in terms of acoustic distance, and even perceptual metrics in some cases, it remains unclear precisely how a less-than-native-sounding L2 French /y/ is produced such that a slight acoustic mismatch might result.

A study could be carried out to test for reuse of L1 motor programs during learning of articulatory strategies for a new speech sound, effectively resituating a novel task (which is not already readily identifiable with some L1 phoneme) within a speaker's L1 task space, within the duration of an experiment. In general, it is expected that L1 articulatory habit will play an outsize role in determining articulatory strategy in L2 phonological categories normally produced outside of L1 articulatory habit, particularly for new learners but possibly for more seasoned veterans of foreign language pedagogy as well. An obviously attractive test case is the frequently studied set of front rounded vowels found in French (Flege and Hillenbrand, 1984; Levy, 2009), which are not only outside of American English articulatory habit in their coordination of tongue front-raising and lip rounding but also in the particulars of their tongue position: French /y/, for instance, is typically produced with a lower tongue dorsum than a typical /i/ and more lip protrusion compared to an American English or French /u/ (Raphael et al., 1979; Wood, 1986).

The benefits for the L2 learner, or even the learner of a single novel contrast, of discovering a strategy based on L1 reuse are clear. If a suitable articulatory strategy can be found within the L1 task space, it may be possible to sound native-like in salient aspects of L2 phonetics (or simply to consistently reproduce something resembling the original contrast) without undertaking a more laborious discovery process. The resulting strategy may in fact be as consistently successful—or more so—than one developed outside the L1 task space.

This pattern could be sought by examining the "endpoint," in articulatory terms, of L2 acquisition in students of an L2, selected from a relatively controlled classroom environment and tested at around the same time (i.e., at the end of the first year of instruction). Such a finding would carry the implication that so-called "fossilization" or non-attainment in spite of continued exposure to native-like linguistic inputs (Han, 2004) may be due to an initial reuse of L1 strategy that is cemented in place over time. This strategy be sufficient for the task of comprehensible L2 production initially, but in the end proves to be suboptimal (and detectable as "accent" by native listeners).

5.3.3 Characterizing the learning curve for L2 articulations

Articulatory reuse could also be revealed in the laboratory by teaching speakers to produce a single "L2" sound. This opens up the possibility of directly characterizing the entirety of the learning curve, which may yield its own insights. Recall that the fast-moving, explicit type of learning I associate with articulatory reuse in Chapter 2 has a characteristic learning curve quite distinct from other types of (implicit) learning: the exploratoration of the task space (Haith and Krakauer, 2013; McDougle et al., 2016) resembles a random walk toward success (Loeb, 2012). This is only especially clear when individual trajectories of learning are examined, since averaging across group members may artificially give the impression of gradual, monotonic improvement where idiosyncratic trial-to-trial variability predominates (Gallistel et al., 2004).

Quantifying acoustic trial-to-trial variability and the number of distinct articulatory strategies attempted during learning of a novel speech sound may thus yield insight into the nature of the speech motor learning process. Some insight may also be gained into whether speech motor learning involves articulatory reuse to the degree argued for here.

5.4 Conclusion

In this thesis, articulatory uniformity is modeled as arising from basic characteristics of motor control and learning; basic characteristics of language acquisition; and motor control as it likely develops during language acquisition. The model advanced here proposes that uniformity in adult phonological inventories arises as a consequence of articulatory reuse during consolidation of speech motor programs; this occurs mainly during L1 acquisition but may also occur in L2 learning. In either situation, articulatory reuse results straightforwardly from a bias towards exploitation of controls that the learner has already mastered, which ensures relatively easy learning of controls that can be applied to new skills.

The model of articulatory uniformity via articulatory reuse predicts that sounds with similar acoustic targets should exhibit a strong tendency toward having similar articulatory implementations of that goal. This prediction is largely confirmed in a study of the so-called *fricative vowels* in Sūzhōu Chinese, which exhibit fricative noise during their production that tends to greatly resemble the fricative noise of a speaker's alveolopalatal fricative and affricate consonants. An individual Sūzhōu Chinese speaker's fricative vowels were found to use lingual articulatory strategies similar to those employed for articulation of their fricative consonants.

In the Sūzhōu Chinese experiment, it is also observed that an individual speaker's choice of a uniform articulatory strategy is not systematically related to their age. One possible explanation of this effect is that the acoustic correlates of uniform articulatory strategies in the Sūzhōu Chinese fricative vowels may not be robustly perceptible or attended to by Sūzhōu Chinese learners. Speakers may attend primarily to fricative noise, which does not appear to vary in easily perceptible ways from speaker to speaker; any difference in formant frequencies between uniform and non-uniform strategies may thus be perceptually unimportant. This suggests that variability in the Sūzhōu Chinese fricative vowels can be added to the known cases of covert, apparently unconstrained variability in speech production. The persistence of this variation actually strengthens the case for uniformity arising from a speaker-side factor: there is no possible source for the uniformity in the learner's ambient language environment, so any uniformity of strategy is most likely driven by speaker-side factors alone.

These speaker-side factors, which I have posited as emerging from the dynamics of learning, merit a more detailed investigation in future work. For instance, it is an open question which aspect of speech production outputs is typically uniformized, if any, in a given situation: articulatory outputs, acoustic outputs (regardless of the articulatory strategies used to achieve them), or both outputs in some combination? Individual variability in prioritizing different sorts of uniformity may enrich our notion of the phonetic grammar with these particulars of phonetic implementation. A deeper examination of the role of articulatory reuse in the development of speech motor repertoires in child learners is also called for. Further work on these and other issues relating to articulatory uniformity and reuse may improve understanding of the substance that makes up phonological and phonetic inventories, how this substance is learned, and the abstract relations among parts of these inventories that speakers maintain through development and into adult speech production.

Bibliography

Abler, William L. (1989). “On the particulate principle of self-diversifying systems”. In: *Journal of Social and Biological Structures* 12.1, pp. 1–13.

Ao, Benjamin Xiaoping (1993). “Phonetics and phonology of Nantong Chinese”. PhD thesis. The Ohio State University.

Articulate Instruments Ltd. (2008). *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*.

Asyik, Abdul Gani (1982). “The agreement system in Acehnese”. In: *Mon-Khmer Studies* 11, pp. 1–33.

Atal, Bishnu S., Jih Jie Chang, Max V. Mathews, and John W. Tukey (1978). “Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique”. In: *The Journal of the Acoustical Society of America* 63.5, pp. 1535–1555.

Ballard, William L. (1969). “Phonological History of Wu”. PhD thesis. University of California, Berkeley.

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Fabian Scheipl, Gabor Grothendieck, and Peter Green (2018). *Linear Mixed-Effects Models using ‘Eigen’ and S4*. R Package, v. 1.1.17, accessed 6 May 2018.

Becker, Michael and Anne-Michelle Tessier (2011). “Trajectories of faithfulness in child-specific phonology”. In: *Phonology* 28.2, pp. 163–196.

Beckman, Mary E. and Jan Edwards (2000). “Lexical frequency effects on young children’s imitative productions”. In: *Papers in Laboratory Phonology V*, pp. 208–218.

Beddor, Patrice Speeter (2009). “A coarticulatory path to sound change”. In: *Language* 85.4, pp. 785–821.

Bender, Mark (2003). *Plum and Bamboo: China’s Suzhou Chantefable Tradition.* University of Illinois Press.

Benson, Bryan L., Joaquin A. Anguera, and Rachael D. Seidler (2011). “A spatial explicit strategy reduces error but interferes with sensorimotor adaptation”. In: *Journal of Neurophysiology* 105.6, pp. 2843–2851.

Berger, Denise J., Reinhard Gentner, Timothy Edmunds, Dinesh K. Pai, and Andrea d’Avella (2013). “Differences in adaptation rates after virtual surgeries provide direct evidence for modularity”. In: *Journal of Neuroscience* 33.30, pp. 12384–12394.

Bernstein, Nikolaj A. (1967). *The co-ordination and regulation of movements.* Pergamon Press.

Berrah, Ahmed-Reda and Rafael Laboissière (1997). “Phonetic code emergence in a society of speech robots: explaining vowel systems and the MUAF Principle”. In: *EuroSpeech97*, pp. 2395–2398.

Best, Catherine T. (1994). “The emergence of native-language phonological influences in infants: A perceptual assimilation model”. In: *The development of speech perception: The transition from speech sounds to spoken words.* Ed. by Judith Goodman and Howard C. Nusbaum. MIT Press, pp. 233–277.

Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell (2001). “Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system”. In: *The Journal of the Acoustical Society of America* 109.2, pp. 775–794.

Best, Catherine T., Gerald W. McRoberts, and Nomathemba M. Sithole (1988). “Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants”. In: *Journal of Experimental Psychology: Human Perception and Performance* 14.3, p. 345.

Birdsong, David (2007). “Nativelike pronunciation among late learners of French as a second language”. In: *Language experience in second language speech learning.* Ed. by Ocke-Schwen Bohn and Murray J. Munro. John Benjamins, pp. 99–116.

Björsten, Sven and Olle Engstrand (1999). “Swedish ‘Damped’ /i/ and /y/: Experimental and Typological Observations”. In: *Proceedings of ICPhS 14*, pp. 1957–1960.

Blevins, Juliette (2004). *Evolutionary phonology: The emergence of sound patterns.* Cambridge University Press.

Boersma, Paul (1993). “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound”. In: *Proceedings of the Institute of Phonetic Sciences.* Vol. 17, pp. 97–110.

Boersma, Paul and David Weenink (2017). *Praat, a system for doing phonetics by computer.* Computer program. Version 5.3.13, retrieved 12 April 2012 from `http://www.praat.org/`.

Bond, Krista M. and Jordan A. Taylor (2015). “Flexible explicit but rigid implicit learning in a visuomotor adaptation task”. In: *Journal of Neurophysiology* 113.10, pp. 3836–3849.

Bouchard, Kristofer E., Nima Mesgarani, Keith Johnson, and Edward F. Chang (2013). “Functional organization of human sensorimotor cortex for speech articulation”. In: *Nature* 495.7441, p. 327.

Braun, Daniel A., Ad Aertsen, Daniel M. Wolpert, and Carsten Mehring (2009). “Motor task variation induces structural learning”. In: *Current Biology* 19.4, pp. 352–357.

Browman, Catherine P. and Louis Goldstein (1989). “Articulatory gestures as phonological units”. In: *Phonology* 6.2, pp. 201–251.

Browman, Catherine P. and Louis Goldstein (1992). “Articulatory phonology: An overview”. In: *Phonetica* 49.3-4, pp. 155–80.

Bruce, Gösta (2010). *Vår fonetiska geografi.* Studentlitteratur AB.

Byrd, Dani and Elliot Saltzman (2003). “The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening”. In: *Journal of Phonetics* 31.2, pp. 149–180.

Cai, Shanqing, Satrajit S. Ghosh, Frank H. Guenther, and Joseph S. Perkell (2010). “Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization”. In: *The Journal of the Acoustical Society of America* 128.4, pp. 2033–2048.

Cao, Jianfen and Ian Maddieson (1992). “An exploration of phonation types in Wu dialects of Chinese”. In: *Journal of Phonetics* 20.1, pp. 77–92.

Carignan, Christopher, Ryan Shosted, Chilin Shih, and Panying Rong (2011). “Compensatory articulation in American English nasalized vowels”. In: *Journal of Phonetics* 39.4, pp. 668–682.

Catford, John C. (1977). *Fundamental Problems in Phonetics*. Midland Books.

Census Office of Jiangsu Province, 江苏省人口普查领导小组办公室 (2012). 江苏省 *2010* 年人口普查资料 *[2010 population census data for Jiangsu province]*. 中国统计出版社 [China Statistics Press].

Census Office of Zhejiang Province, 浙江省人口普查领导小组办公室 (2012). 浙江省 *2010* 年人口普查资料 *[2010 population census data for Zhejiang province]*. 中国统计出版社 [China Statistics Press].

Chao, Yuen Ren (1934). “The non-uniqueness of phonemic solutions of phonetic systems”. In: *Bulletin of the Institute of History and Philosophy, Academia Sinica* 4.4, pp. 363–97.

Chao, Yuen Ren (1967). “Contrastive aspects of the Wu dialects”. In: *Language*, pp. 92–101.

Chen, Matthew Y. (1976). “From middle Chinese to modern Peking”. In: *Journal of Chinese Linguistics* 4.2/3, pp. 113–277.

Chen, Yiya (2008). “The acoustic realization of vowels of Shanghai Chinese”. In: *Journal of Phonetics* 36.4, pp. 629–648.

Chen, Yiya and Carlos Gussenhoven (2015). “Shanghai Chinese”. In: *Journal of the International Phonetic Association* 45.3, pp. 321–337.

Chen, Zhongmin (2014). “On the relationship between tones and initials of the dialects in the Shanghai area”. In: *Proceedings of the Fourth International Symposium on Tonal Aspects of Languages.*

Chen, Zongzhen 陈宗振 (1986). “论西部裕固语的带擦元音 [On Western Yugur’s friction-bearing vowels]”. In: 民族语文 *[Minzu Yuwen]* 2.1.

Chirkova, Katia and Yiya Chen (2013). “Lizu”. In: *Journal of the International Phonetic Association* 43.1, pp. 75–86.

Chirkova, Katia, Dehe Wang, Yiya Chen, Angélique Amelot, and Tanja K. Antolík (2015). “Ersu”. In: *Journal of the International Phonetic Association* 45.2, pp. 187–211.

Cho, Taehong and Peter Ladefoged (1999). “Variation and universals in VOT: Evidence from 18 languages”. In: *Journal of Phonetics* 27.2, pp. 207–229.

Chodroff, Eleanor (2017). “Structured Variation in Obstruent Production and Perception”. PhD thesis. Johns Hopkins University.

Chodroff, Eleanor and Colin Wilson (2017). “Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English”. In: *Journal of Phonetics* 61, pp. 30–47.

Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English.* ERIC.

Clements, George N. (2003). “Feature economy in sound systems”. In: *Phonology* 20.3, pp. 287–333.

Connell, Bruce (2007). “Mambila fricative vowels and Bantu spirantisation”. In: *Africana Linguistica* 13, pp. 7–31.

D’Ausilio, A., L. Maffongelli, E. Bartoli, M. Campanella, E. Ferrari, J. Berry, and L. Fadiga (2014). “Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging”. In: *Phil. Trans. R. Soc. B* 369.1644, p. 20130418.

d’Avella, Andrea and Emilio Bizzi (2005). “Shared and specific muscle synergies in natural motor behaviors”. In: *Proceedings of the National Academy of Sciences* 102.8, pp. 3076–3081.

d’Avella, Andrea, Philippe Saltiel, and Emilio Bizzi (2003). “Combinations of muscle synergies in the construction of a natural motor behavior”. In: *Nature Neuroscience* 6.3, p. 300.

Dart, Sarah N. (1998). “Comparing French and English coronal consonant articulation”. In: *Journal of Phonetics* 26.1, pp. 71–94.

Davidson, Lisa (2006). “Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance”. In: *The Journal of the Acoustical Society of America* 120.1, pp. 407–415.

Davidson, Lisa (2016). “Variability in the implementation of voicing in American English obstruents”. In: *Journal of Phonetics* 54, pp. 35–50.

de Boer, Bart (2000). “Self-organization in vowel systems”. In: *Journal of Phonetics* 28.4, pp. 441–465.

Dell, François (1981). “La langue bai: phonologie et lexique”. In: *Cahiers de linguistique–Asie orientale* 10.1, pp. 143–143.

Dell, François (1994). “Consonnes à prolongement syllabique en Chine”. In: *Cahiers de linguistique—Asie orientale* 23.1.

Diedrichsen, Jörn and Katja Kornysheva (2015). “Motor skill learning between selection and execution”. In: *Trends in Cognitive Sciences* 19.4, pp. 227–233.

Diedrichsen, Jörn, Olivier White, Darren Newman, and Níall Lally (2010). “Use-dependent and error-based learning of motor behaviors”. In: *Journal of Neuroscience* 30.15, pp. 5159–5166.

Docherty, Gerard J. (1992). *The timing of voicing in British English obstruents.* Netherlands Phonetic Archives 9. Walter de Gruyter.

Doke, Clement M. (1963). *Textbook of Zulu grammar*. Longmans Southern Africa.

Duanmu, San (2007). *Phonology of Standard Chinese.* Oxford University Press.

Dwyer, Arienne M. (1998). “The texture of tongues: Languages and power in China”. In: *Nationalism and Ethnic Politics* 4.1-2, pp. 68–85.

Dwyer, Arienne M. (2008). “Tonogenesis in Southeastern Monguor”. In: *Lessons from Documented Endangered Languages*. Ed. by Keith Harrison and David Rood. John Benjamins, pp. 111–128.

Easton, Thomas A. (1972). “On the normal use of reflexes”. In: *American Scientist* 60.5, pp. 591–599.

Engstrand, Olle, Sven Björsten, Björn Lindblom, Gösta Bruce, and Anders Eriksson (2000). “Hur udda är Viby-i? Experimentella och typologiska observationer”. In: *Folkmålsstudier* 39, pp. 83–95.

Esling, John H. (2005). “There Are No Back Vowels: The Laryngeal Articulator Model”. In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 50.1–4, pp. 13–44.

Fagyal, Zsuzsanna and Christine Moisset (1999). “Sound change and articulatory release: where and why are high vowels devoiced in Parisian French?” In: *Proceedings of ICPhS 14*, pp. 309–312.

Fant, Gunnar (1960). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations.* Walter de Gruyter.

Faytak, Matthew (2014). “High vowel fricativization and chain shift”. In: *UC Berkeley Phonology Lab Annual Report*. Vol. 10.

Faytak, Matthew (2017). “Sonority in some languages of the Cameroon Grassfields”. In: *Challenging Sonority*. Ed. by Martin J. Ball and Nicole Müller. Equinox.

Faytak, Matthew and Susan Lin (2015). “Articulatory variability and fricative noise in Standard Mandarin apical vowels”. In: *Proceedings of ICPhS 18*.

Ferguson, Charles A. and Carol B. Farwell (1975). “Words and sounds in early language acquisition”. In: *Language*, pp. 419–439.

Field, Keith L. (1997). “A grammatical overview of Santa Mongolian”. PhD thesis. University of California, Santa Barbara.

Fiore, Lynne E. (1987). *A phonology of Limbum*. Univ. of Texas, Arlington.

Flege, James E. (1982). “Laryngeal timing and phonation onset in utterance-initial English stops”. In: *Journal of Phonetics* 10, pp. 177–92.

Flege, James E., Ocke-Schwen Bohn, and Sunyoung Jang (1997). “Effects of experience on non-native speakers’ production and perception of English vowels”. In: *Journal of Phonetics* 25.4, pp. 437–470.

Flege, James E. and James Hillenbrand (1984). “Limits on phonetic accuracy in foreign language speech production”. In: *The Journal of the Acoustical Society of America* 76.3, pp. 708–721.

Flege, James and W. M. Brown (1982). “Effects of utterance position on English speech timing”. In: *Phonetica* 39.6, pp. 337–357.

Flemming, Edward (2004). “Contrast and perceptual distinctiveness”. In: *Phonetically based phonology*. Ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade. Cambridge University Press, pp. 232–276.

Fowler, Carol A. (1986). “An event approach to the study of speech perception from a direct-realist perspective”. In: *Status report on speech research* 14, p. 139.

Fowler, Carol A. and Michael T. Turvey (1980). “Immediate compensation in bite-block speech”. In: *Phonetica* 37.5-6, pp. 306–326.

Fox, John, Sanford Weisberg, Brad Price, Daniel Adler, Douglas Bates, Gabriel Baud-Bovy, Ben Bolker, Steve Ellison, David Firth, Michael Friendly, Gregor Gorjanc, Spencer Graves, Richard Heiberger, Rafael Laboissière, Martin Maechler, Georges Monette, Duncan Murdoch, Henric Nilsson, Derek Ogle, Brian Ripley, William Venables, Steve Walker, David Winsemius, and Achim Zeileis (2018). *Companion to Applied Regression.* R Package, v. 3.0.0, accessed 6 May 2018.

Friston, Karl (2011). “What is optimal about motor control?” In: *Neuron* 72.3, pp. 488–498.

Fruehwald, Josef (2013). “Phonological involvement in phonetic change”. PhD thesis.

Gallistel, Charles R., Stephen Fairhurst, and Peter Balsam (2004). “The learning curve: implications of a quantitative analysis”. In: *Proceedings of the National Academy of Sciences* 101.36, pp. 13124–13131.

Garrett, Andrew and Keith Johnson (2013). “Phonetic bias in sound change”. In: *Origins of Sound Patterns: Approaches to Phonologization.* Ed. by Alan C. L. Yu. Oxford.

Gentil, Michèle (1992). “Variability of motor strategies”. In: *Brain and Language* 42.1, pp. 30–37.

Gick, Bryan and Donald Derrick (2009). “Aero-tactile integration in speech perception”. In: *Nature* 462.7272, p. 502.

Gick, Bryan and Ian Stavness (2013). “Modularizing speech”. In: *Frontiers in Psychology* 4, p. 977.

Goldstein, Louis, Dani Byrd, and Elliot Saltzman (2006). “The role of vocal tract gestural action units in understanding the evolution of phonology”. In: *Action to language via the mirror neuron system.* Ed. by Michael A Arbib, pp. 215–49.

Goldstein, Louis and Carol A. Fowler (2003). “Articulatory phonology: A phonology for public language use”. In: *Phonetics and phonology in language comprehension and production: Differences and similarities.* Ed. by Niels O Schiller and Antje S Meyer, pp. 159–207.

Goodman, Leo A. (1961). “Snowball sampling”. In: *The Annals of Mathematical Statistics* 32.1, pp. 148–170.

Gu, Chong (2002). *Smoothing spline ANOVA models.* Springer Science & Business Media.

Guenther, Frank H. (1994). “A neural network model of speech acquisition and motor equivalent speech production”. In: *Biological Cybernetics* 72.1, pp. 43–53.

Guenther, Frank H. (1995). “Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production.” In: *Psychological Review* 102.3, p. 594.

Guenther, Frank H., Carol Y. Espy-Wilson, Suzanne E. Boyce, Melanie L. Matthies, Majid Zandipour, and Joseph S. Perkell (1999). “Articulatory tradeoffs reduce acoustic variability during American English /r/ production”. In: *The Journal of the Acoustical Society of America* 105.5, pp. 2854–2865.

Guenther, Frank H., Satrajit S. Ghosh, and Jason A. Tourville (2006). “Neural modeling and imaging of the cortical interactions underlying syllable production”. In: *Brain and Language* 96.3, pp. 280–301.

Guigon, Emmanuel, Pierre Baraduc, and Michel Desmurget (2008). “Optimality, stochasticity, and variability in motor behavior”. In: *Journal of Computational Neuroscience* 24.1, pp. 57–68.

Haith, Adrian M. and John W. Krakauer (2013). “Model-based and model-free mechanisms of human motor learning”. In: *Progress in Motor Control*. Ed. by Michael Richardson, Michael Riley, and Kevin Shockley. Vol. 782. Advances in Experimental Medicine and Biology. Springer, pp. 1–21.

Han, ZhaoHong (2004). *Fossilization in Adult Second Language Acquisition*. Multilingual Matters.

Hickok, Gregory (2014). “The architecture of speech production and the role of the phoneme in speech processing”. In: *Language, Cognition and Neuroscience* 29.1, pp. 2–20.

Hickok, Gregory, John Houde, and Feng Rong (2011). “Sensorimotor integration in speech processing: computational basis and neural organization”. In: *Neuron* 69.3, pp. 407–422.

Honda, Kiyoshi (1996). “Organization of tongue articulation for vowels”. In: *Journal of Phonetics* 24.1, pp. 39–52.

Hou, Chao 侯超 (2009). “合肥方言带擦元音 [ɿ] 的实验研究 [An experimental study on [ɿ] in Hefei Dialect”. In: 南京师范大学文学院学报 *[Journal of School of Chinese Language and Culture, Nanjing Normal University]* 1, pp. 172–77.

Houde, John F. and Michael I. Jordan (1998). “Sensorimotor adaptation in speech production”. In: *Science* 279.5354, pp. 1213–1216.

Houde, John F. and Srikantan S. Nagarajan (2011). “Speech production as state feedback control”. In: *Frontiers in Human Neuroscience* 5.

Howie, John Marshall (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge University Press.

Hu, Fang 胡方 (2007). “论宁波方言和苏州方言前高元音的区别特征—兼谈高元音继续高化现象 [On the distinctive features of the high front vowels in Ningbo and Suzhou Wu dialects with reference to sound changes of high vowels]”. In: 中国语文 *[Zhongguo Yuwen]* 5, pp. 455–65.

Huberdeau, David M., John W. Krakauer, and Adrian M. Haith (2015). “Dual-process decomposition in human sensorimotor adaptation”. In: *Current Opinion in Neurobiology* 33, pp. 71–77.

Hueber, Thomas, Guido Aversano, G. Cholle, Bruce Denby, Gérard Dreyfus, Yacine Oussar, Pierre Roussel, and Maureen Stone (2007). “Eigentongue feature extraction for an ultrasound-based silent speech interface”. In: *ICASSP 2007*. Vol. 1. IEEE, pp. I–1245.

Hyman, Larry M. (1980). “Babanki and the Ring group”. In: *Les Classes Nominales dans le Bantou des Grassfields* 1, pp. 225–258.

Hyman, Larry M. (1981). *Noni grammatical structure: with special reference to verb morphology*. University of Southern California.

Jeanne, LaVerne Masayesva (1978). “Aspects of Hopi grammar”. PhD thesis. Massachusetts Institute of Technology.

Jiangxi Bureau of Statistics, 江西省统计局 (2012). 江西省 *2010* 年人口普查资料 *[2010 population census data for Jiangxi province]*. 中国统计出版社 [China Statistics Press].

Johnson, Keith (2006). “Resonance in an exemplar-based lexicon: The emergence of social identity and phonology”. In: *Journal of Phonetics* 34.4, pp. 485–499.

Johnson, Keith, Peter Ladefoged, and Mona Lindau (1993). “Individual differences in vowel production”. In: *The Journal of the Acoustical Society of America* 94.2, pp. 701–714.

Jones, Daniel (1956). *An Outline of English Phonetics.* Heffer.

Kager, René (1999). *Optimality Theory.* Cambridge University Press.

Kak, Avinash C. and Azriel Rosenfeld (1982). *Digital picture processing.* Academic Press.

Karlgren, Bernhard (1926). *Études sur la phonologie chinoise.* EJ Brill.

Keating, Patricia (1984). “Phonetic and phonological representation of stop consonant voicing”. In: *Language* 60.2, pp. 286–319.

Keating, Patricia (1990). “Phonetic representations in a generative grammar”. In: *Journal of Phonetics* 18.3, pp. 321–334.

Keating, Patricia (2003). “Phonetic and other influences on voicing contrasts”. In: *Proceedings of ICPhS 15*, pp. 375–378.

Kelso, J. Scott, Betty Tuller, Eric Vatikiotis-Bateson, and Carol A. Fowler (1984). “Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures.” In: *Journal of Experimental Psychology: Human Perception and Performance* 10.6, p. 812.

Kim, S.S. (2003). “Santa”. In: *The Mongolic Languages.* Ed. by Juha Janhunen. Routledge, pp. 346–363.

Kirby, James P. (2013). “The role of probabilistic enhancement in phonologization”. In: *Origins of Sound Change: Approaches to Phonologization, Oxford, UK: Oxford University Press* 22, pp. 228–246.

Ko, S.-J. and Yong Hoon Lee (1991). “Center weighted median filters and their applications to image enhancement”. In: *IEEE transactions on circuits and systems* 38.9, pp. 984–993.

Koenig, Laura L., Christine H. Shadle, Jonathan L. Preston, and Christine R. Mooshammer (2013). “Toward improved spectral measures of /s/: Results from adolescents”. In: *Journal of Speech, Language, and Hearing Research* 56.4, pp. 1175–1189.

Krakow, Rena A., Patrice S. Beddor, Louis M. Goldstein, and Carol A. Fowler (1988). “Coarticulatory influences on the perceived height of nasal vowels”. In: *The Journal of the Acoustical Society of America* 83.3, pp. 1146–1158.

Ladefoged, Peter and Mona Lindau (1989). “Modeling articulatory-acoustics relations: A comment on Stevens’ “On the quantal nature of speech””. In: *Journal of Phonetics* 17.1-2, pp. 99–106.

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

Lee-Kim, Sang-im (2014). “Revisiting Mandarin ‘apical vowels’: An articulatory and acoustic study”. In: *Journal of the International Phonetic Association* 44.3, pp. 261–282.

Lee, Wai-Sum and Eric Zee (2015). “Apical vowels”. In: *Encyclopedia of Chinese Language and Linguistics.* Ed. by Rint Sybesma. Accessed online 01 September 2017.

Levy, Erika S. (2009). “Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French”. In: *The Journal of the Acoustical Society of America* 125.2, p. 1138.

Levy, Erika S. and Franzo F. Law (2010). “Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship”. In: *The Journal of the Acoustical Society of America* 128.3, pp. 1290–1305.

Li, Fang-Kuei (1966). “The zero initial and the zero syllabic”. In: *Language*, pp. 300–302.

Li, Min 李民 and Ming 马明 Ma (1983). 凉山彝语语音概论. Introduction to the sounds of the Liangshan Yi language. 四川民族出版社 [Sichuan Nationalities Press].

Li, Min, Chandra Kambhamettu, and Maureen Stone (2005). “Automatic contour tracking in ultrasound images”. In: *Clinical Linguistics & Phonetics* 19.6-7, pp. 545–554.

Liljencrants, Johan and Björn Lindblom (1972). “Numerical simulation of vowel quality systems: The role of perceptual contrast”. In: *Language*, pp. 839–862.

Lindau, Mona (1978). “Vowel features”. In: *Language* 54.3, pp. 541–563.

Lindblom, Björn (1984). “Can the models of evolutionary biology be applied to phonetic problems?” In: *Proceedings of ICPhS 10*, pp. 67–81.

Lindblom, Björn (1990). “Explaining phonetic variation: A sketch of the H&H theory”. In: *Speech production and speech modelling*. Ed. by W. Hardcastle and A. Marchal. Springer, pp. 403–439.

Lindblom, Björn (1998). “Systemic constraints and adaptive change in the formation of sound structure”. In: *Approaches to the evolution of language: Social and cognitive bases*. Ed. by James R. Hurford, Michael Studdert-Kennedy, and Chris Knight. Cambridge University Press Cambridge, pp. 242–264.

Lindblom, Björn, Randy Diehl, Sang-Hoon Park, and Giampiero Salvi (2011). “Sound systems are shaped by their users: The recombination of phonetic substance”. In: *Where do phonological features come from?* Ed. by Nick G. Clements and Rachid Ridouane. John Benjamins, pp. 67–97.

Lindblom, Björn and Johan Sundberg (1971). “Acoustical consequences of lip, tongue, jaw, and larynx movement”. In: *The Journal of the Acoustical Society of America* 50.4B, pp. 1166–1179.

Ling, Feng (2007). “The articulatory and acoustic study of fricative vowels in Suzhou Chinese”. In: *Proceedings of ICPhS 16*.

Ling, Feng 凌峰 (2009). “A phonetic study of the vowel system in Suzhou Chinese”. PhD thesis. City University of Hong Kong.

Linker, Wendy J. (1982). “Articulatory and acoustic correlates of labial activity: a cross-linguistic study”. PhD thesis. University of California, Los Angeles.

Locke, John L. (1983). *Phonological acquisition and change*. Academic Press.

Loeb, Gerald E (2012). “Optimal isn’t good enough”. In: *Biological Cybernetics* 106.11-12, pp. 757–765.

Loeb, Gerald E. (1983). “Finding common ground between robotics and physiology”. In: *Trends in Neurosciences* 6, pp. 203–204.

Löfqvist, Anders and Hirohide Yoshioka (1981). “Interarticulator programming in obstruent production”. In: *Phonetica* 38.1-3, pp. 21–34.

Mackie, Scott and Jeff Mielke (2011). “Feature economy in natural, random, and synthetic inventories”. In: *Where do phonological features come from?* Ed. by Nick G. Clements and Rachid Ridouane. John Benjamins, pp. 43–63.

Maddieson, Ian (1996). “Gestural economy”. In: *UCLA Working Papers in Phonetics*, pp. 1–6.

Maddieson, Ian and Sandra Ferrari Disner (1984). *Patterns of Sounds.* Cambridge University Press.

Maeda, Shinji (1990). “Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model”. In: *Speech Production and Speech Modelling.* Ed. by Hardcastle William J. and Alain Marchal. Springer, pp. 131–149.

Maekawa, Kikuo and Hideaki Kikuchi (2005). “Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report”. In: *Voicing in Japanese.* Ed. by Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara. Vol. 84. Walter de Gruyter, p. 205.

Malfait, Nicole, Paul L. Gribble, and David J. Ostry (2005). “Generalization of motor learning based on multiple field exposures and local adaptation”. In: *Journal of Neurophysiology* 93.6, pp. 3327–3338.

Maniwa, Kazumi, Allard Jongman, and Travis Wade (2009). “Acoustic characteristics of clearly spoken English fricatives”. In: *The Journal of the Acoustical Society of America* 125.6, pp. 3962–3973.

Manto, Mario, James M. Bower, Adriana Bastos Conforto, José M. Delgado-García, Suzete Nascimento Farias da Guarda, Marcus Gerwig, Christophe Habas, Nobuhiro Hagura, Richard B. Ivry, Peter Mariën, et al. (2012). “Consensus paper: roles of the cerebellum in motor control—the diversity of ideas on cerebellar involvement in movement”. In: *The Cerebellum* 11.2, pp. 457–487.

Martinet, André (1955). *Économie des changements phonétiques: traité de phonologie diachronique.* Vol. 10. Éditions A. Francke.

Martinet, André (1968). “Phonetics and linguistic evolution”. In: *Manual of phonetics, revised and extended edition.* Ed. by Bertil Malmberg. North-Holland Publishing, pp. 464–487.

Mazzoni, Pietro and John W. Krakauer (2006). “An implicit plan overrides an explicit strategy during visuomotor adaptation”. In: *Journal of Neuroscience* 26.14, pp. 3642–3645.

McAllister Byun, Tara, Sharon Inkelas, and Yvan Rose (2016). “The A-map model: Articulatory reliability in child-specific phonology”. In: *Language* 92.1, pp. 141–178.

McDougle, Samuel D., Krista M. Bond, and Jordan A. Taylor (2015). “Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning”. In: *Journal of Neuroscience* 35.26, pp. 9568–9579.

McDougle, Samuel D., Richard B. Ivry, and Jordan A. Taylor (2016). “Taking aim at the cognitive side of learning in sensorimotor adaptation tasks”. In: *Trends in Cognitive Sciences* 20.7, pp. 535–544.

McFarland, David H., Shari R. Baum, and Caroline Chabot (1996). “Speech compensation to structural modifications of the oral cavity”. In: *The Journal of the Acoustical Society of America* 100.2, pp. 1093–1104.

McReynolds, Leija V. and Shannon Bennett (1972). “Distinctive feature generalization in articulation training”. In: *Journal of Speech and Hearing Disorders* 37.4, pp. 462–470.

Medjo Mvé, Pithier (1997). “Essai sur la phonologie panchronique des parlers fang du Gabon et ses implications historiques”. PhD thesis. Univ. Lumière Lyon II.

Ménard, Lucie, Jean-Luc Schwartz, and Jérôme Aubin (2008). “Invariance and variability in the production of the height feature in French vowels”. In: *Speech Communication* 50.1, pp. 14–28.

Menn, Lise (1978). “Phonological units in beginning speech”. In: *Syllables and Segments.* Ed. by Alan Bell and Joan B. Hooper. North-Holland Amsterdam, pp. 157–171.

Menn, Lise (1983). “Development of articulatory, phonetic, and phonological capabilities”. In: *Language Production.* Ed. by Brian Butterworth. Vol. 2, pp. 3–50.

Miall, R. Christopher and Daniel M. Wolpert (1996). “Forward models for physiological motor control”. In: *Neural Networks* 9.8, pp. 1265–1279.

Michaud, Alexis (2008). “Phonemic and tonal analysis of Yongning Na”. In: *Cahiers de linguistique–Asie orientale* 37.2, pp. 159–196.

Michaud, Alexis and Likun 和丽昆 He (2015). “Phonemic and tonal analysis of the Pianding dialect of Naxi (Dadong County, Lijiang Municipality)”. In: *Cahiers de linguistique–Asie orientale* 44.1, pp. 1–35.

Mielke, Jeff (2008). *The Emergence of Distinctive Features.* Oxford University Press.

Mielke, Jeff (2015). “An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons”. In: *The Journal of the Acoustical Society of America* 137.5, pp. 2858–2869.

Mielke, Jeff, Adam Baker, and Diana Archangeli (2016). “Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/”. In: *Language* 92.1, pp. 101–140.

Nasir, Sazzad M. and David J. Ostry (2009). “Auditory plasticity and speech motor learning”. In: *Proceedings of the National Academy of Sciences* 106.48, pp. 20470–20475.

Niziolek, Caroline A., Srikantan S. Nagarajan, and John F. Houde (2013). “What does motor efference copy represent? Evidence from speech production”. In: *Journal of Neuroscience* 33.41, pp. 16110–16116.

Noiray, Aude, Khalil Iskarous, and Douglas H. Whalen (2014). “Variability in English vowels is comparable in articulation and acoustics”. In: *Laboratory Phonology* 5.2, pp. 271–288.

Norman, Jerry (1988). *Chinese.* Cambridge University Press.

Ohala, John J. (1979). “Moderator’s introduction to *Symposium on Phonetic Universals in Phonological Systems and their Explanation*”. In: *Proceedings of ICPhS 3*, pp. 181–185.

Ohala, John J. (1983). “The origin of sound patterns in vocal tract constraints”. In: *The Production of Speech.* Ed. by P.F. MacNeilage. Springer, pp. 189–216.

Ohala, John J. (1989). “Sound change is drawn from a pool of synchronic variation”. In: *Language change: Contributions to the study of its causes.* Ed. by Leiv E. Breivik and Ernst H. Jahr. Mouton de Gruyter, pp. 173–198.

Ohala, John J. (1990). “There is no interface between phonology and phonetics: a personal view”. In: *Journal of Phonetics* 18.2, pp. 153–172.

Ohala, John J. (1993). “The phonetics of sound change”. In: *Historical Linguistics: Problems and Perspectives.* Ed. by C. Jones. Longman, pp. 237–78.

Olson, Kenneth S. and Yohann Meynadier (2015). “On Medumba bilabial trills and vowels”. In: *Proceedings of ICPhS 18.*

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”. In: *The Journal of the Acoustical Society of America* 93.5, pp. 2948–2961.

Perrier, Pascal, David J. Ostry, and Rafael Laboissière (1996). “The equilibrium point hypothesis and its application to speech motor control”. In: *Journal of Speech, Language, and Hearing Research* 39.2, pp. 365–378.

Pierrehumbert, Janet B. (2008). “Word-specific phonetics”. In: *Laboratory Phonology* 4.1, p. 101.

Pile, Elizabeth J. S., H. R. Dajani, D. W. Purcell, and Kevin G. Munhall (2007). “Talking under conditions of altered auditory feedback: does adaptation of one vowel generalize to other vowels”. In: *Proceedings of ICPhS 16*, pp. 645–648.

Prince, Alan and Paul Smolensky (2004). *Optimality Theory: Constraint interaction in generative grammar.* John Wiley & Sons.

Proctor, Michael I., Christine H. Shadle, and Khalil Iskarous (2010). “Pharyngeal articulation in the production of voiced and voiceless fricatives”. In: *The Journal of the Acoustical Society of America* 127.3, pp. 1507–1518.

Pulleyblank, Edwin G. (1984). “Vowelless Chinese? An application of the three tiered theory of syllable structure to Pekingese”. In: *Proceedings of the XVI International Conference on Sino-Tibetan Languages and Linguistics.* Vol. 2, pp. 568–610.

Pullum, Geoffrey K. and William A. Ladusaw (1996). *Phonetic Symbol Guide.* University of Chicago Press.

Qian, Nairong 钱乃荣 (1992). 当代吴语研究 *[Studies in modern Wu languages].* 上海教育出版社 [Shanghai Education Press].

Qumutiexi, 曲木铁西 (2010). 彝语义诺话研究. Yinuo studies. 民族出版社 [Minority Peoples’ Press].

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* v. 3.5.0, accessed 6 May 2018. Vienna, Austria: R Foundation for Statistical Computing.

Ramanarayanan, Vikram, Louis Goldstein, and Shrikanth S. Narayanan (2013). “Spatiotemporal articulatory movement primitives during speech production: Extraction, interpretation, and validation”. In: *The Journal of the Acoustical Society of America* 134.2, pp. 1378–1394.

Ranganathan, Rajiv, Jon Wieser, Kristine M. Mosier, Ferdinando A. Mussa-Ivaldi, and Robert A. Scheidt (2014). “Learning redundant motor tasks with and without overlapping dimensions: facilitation and interference effects”. In: *The Journal of Neuroscience* 34.24, pp. 8289–8299.

Raphael, Lawrence J., Fredericka Bell-Berti, René Collier, and Thomas Baer (1979). “Tongue position in rounded and unrounded front vowel pairs”. In: *Language and Speech* 22.1, pp. 37–48.

Riad, Tomas (2013). *The phonology of Swedish.* Oxford University Press.

Rose, Phil (1982). “An acoustically based phonetic description of the syllable in the Zhenhai dialect”. PhD thesis. University of Cambridge.

Rose, Phil (1989). “Phonetics and phonology of Yang tone; phonation types in Zhenhai”. In: *Cahiers de linguistique—Asie orientale* 18.2, pp. 229–245.

Rugy, Aymar de, Gerald Loeb, and Timothy Carroll (2013). “Are muscle synergies useful for neural control?” In: *Frontiers in computational neuroscience* 7, p. 19.

Saltzman, Elliot L. and Kevin G. Munhall (1989). “A dynamical approach to gestural patterning in speech production”. In: *Ecological Psychology* 1.4, pp. 333–382.

Sancier, Michele L. and Carol A. Fowler (1997). “Gestural drift in a bilingual speaker of Brazilian Portuguese and English”. In: *Journal of Phonetics* 25.4, pp. 421–436.

Sanders, Ira and Liancai Mu (2013). “A Three-Dimensional Atlas of Human Tongue Muscles”. In: *The Anatomical Record* 296.7, pp. 1102–1114.

Sanguineti, Vittorio, Rafael Laboissière, and Yohan Payan (1997). “A control model of human tongue movements in speech”. In: *Biological Cybernetics* 77.1, pp. 11–22.

Schötz, Susanne, Johan Frid, Lars Gustafsson, and Anders Löfqvist (2014). “Tongue articulation dynamics of /iː, yː, ʉː/ in Stockholm, Gothenburg and Malmöhus Swedish”. In: *Proceedings of Fonetik 2014*, pp. 17–22.

Schötz, Susanne, Johan Frid, and Anders Löfqvist (2011). “Exotic vowels in Swedish: a project description and an articulographic and acoustic pilot study of /i/”. In: *Proceedings of Fonetik 2011.* Speech, Music and Hearing, KTH Stockholm.

Schwartz, Jean-Luc, Anahita Basirat, Lucie Ménard, and Marc Sato (2012). “The Perception−for−Action−Control Theory (PACT): A perceptuo−motor theory of speech perception”. In: *Journal of Neurolinguistics* 25.5, pp. 336–354.

Schwartz, Jean-Luc, Louis-Jean Boë, and Christian Abry (2007). “Linking Dispersion-Focalization Theory and the Maximum Utilization of the Available Distinctive Features Principle in a Perception-for-Action-Control Theory”. In: *Experimental Approaches to Phonology.* Ed. by Maria-Josep Solé, Patrice S. Beddor, and Manjari Ohala. Oxford University Press, Oxford, pp. 104–124.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry (1997a). “Major trends in vowel system inventories”. In: *Journal of Phonetics* 25.3, pp. 233–253.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry (1997b). “The dispersion-focalization theory of vowel systems”. In: *Journal of Phonetics* 25.3, pp. 255–286.

Scovel, Thomas (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech.* Newbury House Rowley, MA.

Shadle, Christine H. (1990). “Articulatory-acoustic relationships in fricative consonants”. In: *Speech production and speech modelling.* Ed. by William J. Hardcastle and Alain Marchal. Springer, pp. 187–209.

Shadle, Christine H., Wei-Rong Chen, and Douglas H. Whalen (2016). “Stability of the main resonance frequency of fricatives despite changes in the first spectral moment”. In: *The Journal of the Acoustical Society of America* 140.4, pp. 3219–3220.

Shadmehr, Reza and John W. Krakauer (2008). “A computational neuroanatomy for motor control”. In: *Experimental Brain Research* 185.3, pp. 359–381.

Shadmehr, Reza, Maurice A. Smith, and John W. Krakauer (2010). “Error correction, sensory prediction, and adaptation in motor control”. In: *Annual Review of Neuroscience* 33, pp. 89–108.

Shanghai Bureau of Statistics, 上海市统计局 (2012). 上海市 *2010* 年人口普查资料 *[2010 population census data for Shanghai city].* 中国统计出版社 [China Statistics Press].

Sherard, Michael (1972). “Shanghai Phonology”. PhD thesis. Cornell University.

Simmons, Richard VanNess (1999). *Chinese dialect classification: A comparative approach to Harngjou, Old Jintarn, and Common Northern Wu.* Vol. 188. John Benjamins Publishing.

Simons, Gary F. and Charles D. Fennig (2015). *Ethnologue: Languages of the world.* Vol. 20. Online version: `http://ethnologue.com`, accessed 10 August 2017. SIL international Dallas, TX.

Smith, Caroline L. (2003). “Vowel devoicing in contemporary French”. In: *Journal of French Language Studies* 13.2, pp. 177–194.

Sohn, M. Hongchul and Lena H. Ting (2016). “Suboptimal muscle synergy activation patterns generalize their motor function across postures”. In: *Frontiers in Computational Neuroscience* 10, p. 7.

Stavness, Ian, Bryan Gick, Donald Derrick, and Sidney Fels (2012). “Biomechanical modeling of English /ɹ/ variants”. In: *The Journal of the Acoustical Society of America* 131.5, EL355–EL360.

Stevens, Kenneth N. and Samuel Jay Keyser (1989). “Primary features and their enhancement in consonants”. In: *Language* 65.1, pp. 81–106.

Studdert-Kennedy, Michael (1998). “The particulate origins of language generativity: from syllable to gesture”. In: *Approaches to the evolution of language: Social and cognitive bases.* Ed. by James R. Hurford, Michael Studdert-Kennedy, and Chris Knight. Cambridge University Press Cambridge, pp. 242–264.

Sülzenbrück, Sandra and Herbert Heuer (2009). “Functional independence of explicit and implicit motor adjustments”. In: *Consciousness and Cognition* 18.1, pp. 145–159.

Takano, Sayoko and Kiyoshi Honda (2007). “An MRI analysis of the extrinsic tongue muscles during vowel production”. In: *Speech Communication* 49.1, pp. 49–58.

Taylor, Jordan A. and Richard B. Ivry (2011). “Flexible cognitive strategies during motor learning”. In: *PLoS Computational Biology* 7.3, e1001096.

Taylor, Jordan A., Nola M. Klemfuss, and Richard B. Ivry (2010). “An explicit strategy prevails when the cerebellum fails to compute movement errors”. In: *The Cerebellum* 9.4, pp. 580–586.

Taylor, Jordan A., John W. Krakauer, and Richard B. Ivry (2014). “Explicit and implicit contributions to learning in a sensorimotor adaptation task”. In: *Journal of Neuroscience* 34.8, pp. 3023–3032.

Thibeault, Mélanie, Lucie Ménard, Shari R. Baum, Gabrielle Richard, and David H. McFarland (2011). “Articulatory and acoustic adaptation to palatal perturbation”. In: *The Journal of the Acoustical Society of America* 129.4, pp. 2112–2120.

Tian, Xing and David Poeppel (2010). “Mental imagery of speech and movement implicates the dynamics of internal forward models”. In: *Frontiers in Psychology* 1, p. 166.

Todorov, Emanuel and Michael I. Jordan (2002). “Optimal feedback control as a theory of motor coordination”. In: *Nature Neuroscience* 5.11, p. 1226.

Todorov, Emanuel, Weiwei Li, and Xiuchuan Pan (2005). “From task parameters to motor synergies: A hierarchical framework for approximately optimal control of redundant manipulators”. In: *Journal of Field Robotics* 22.11, pp. 691–710.

Tourville, Jason A. and Frank H. Guenther (2011). “The DIVA model: A neural theory of speech acquisition and production”. In: *Language and Cognitive Processes* 26.7, pp. 952–981.

Traill, Anthony, James S. Khumalo, and Peter Fridjhon (1987). “Depressing facts about Zulu”. In: *African Studies* 46.2, pp. 255–274.

Tremblay, Stéphanie, Guillaume Houle, and David J. Ostry (2008). “Specificity of speech motor learning”. In: *The Journal of Neuroscience* 28.10, pp. 2426–2434.

Turvey, Michael T. (1977). “Preliminaries to a theory of action with reference to vision”. In: *Perceiving, acting and knowing*. Ed. by R. Shaw and J. Bransford, pp. 211–265.

Ueda, Yuichi, Tomoya Hamakawa, Tadashi Sakata, Syota Hario Hario, and Akira Watanabe (2007). “A real-time formant tracker based on the inverse filter control method”. In: *Acoustical Science and Technology of the Acoustical Science of Japan* 28.4, pp. 271–274.

Verhoef, Tessa, Simon Kirby, and Bart de Boer (2014). “Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals”. In: *Journal of Phonetics* 43, pp. 57–68.

Vihman, Marilyn (2014). *Phonological Development: The first two years*. John Wiley & Sons.

Vihman, Marilyn and William Croft (2007). “Phonological development: toward a “radical” templatic phonology”. In: *Linguistics* 45.4, pp. 683–725.

Vihman, Marilyn, Rory A. DePaolis, and Tamar Keren-Portnoy (2009). “A dynamic systems approach to babbling and words”. In: *The Cambridge Handbook of Child Language*. Ed. by Edith L. Bavin. Cambridge University Press, pp. 163–184.

Vihman, Marilyn and Shelley L. Velleman (2000). “Phonetics and the origins of phonology”. In: *Phonological knowledge: Its nature and status*. Ed. by Noel Burton-Roberts, Philip Carr, and Gerard Docherty. Oxford University Press, pp. 305–339.

Villacorta, Virgilio M., Joseph S. Perkell, and Frank H. Guenther (2007). “Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception”. In: *The Journal of the Acoustical Society of America* 122.4, pp. 2306–2319.

Vydrine, Valentin (2004). “Areal and genetic features in West Mande and South Mande phonology: In what sense did Mande languages evolve?” In: *Journal of West African Languages* 30.2, pp. 113–126.

Wang, Ping 汪平 (1987). “苏州音系再分析 [Suzhou phonology reanalyzed]”. In: 语言研究 *[Studies in Language and Linguistics]* 1, pp. 41–48.

Wang, Shuangcheng 王双成 (2006). “青海方言元音 [i] 的舌尖化音变 [Apicalization sound change of the Qinghai dialects’ [i] vowel]”. In: 中国语文 *[Zhongguo Yuwen]* 4, pp. 359–363.

Wang, Shuangcheng 王双成 (2010). “安多藏语 i 的舌尖化及其类型学意义 [Sound Change of [i] in Amdo Tibetan and its Typological Meaning]”. In: 语言研究 *[Studies in Language and Linguistics]* 30.2, pp. 122–27.

Westbury, John R., Michiko Hashi, and Mary J. Lindstrom (1998). “Differences among speakers in lingual articulation for American English /ɹ/”. In: *Speech Communication* 26.3, pp. 203–226.

Westerberg, Fabienne (2016). “An auditory, acoustic, articulatory, and sociophonetic study of Swedish Viby-i”. MA thesis. University of Glasgow.

Wiese, Richard (1997). “Underspecification and the description of Chinese vowels”. In: *Studies in Chinese phonology*. Ed. by Jialing Wang and Norval Smith. Walter de Gruyter, pp. 219–250.

Wilson, Robert C., Andra Geana, John M. White, Elliot A. Ludvig, and Jonathan D. Cohen (2014). “Humans use directed and random exploration to solve the explore–exploit dilemma.” In: *Journal of Experimental Psychology: General* 143.6, p. 2074.

Wolpert, Daniel M. (1997). “Computational approaches to motor control”. In: *Trends in Cognitive Sciences* 1.6, pp. 209–216.

Wolpert, Daniel M. and J. Randall Flanagan (2001). “Motor prediction”. In: *Current Biology* 11.18, R729–R732.

Wolpert, Daniel M., Zoubin Ghahramani, and J. Randall Flanagan (2001). “Perspectives and problems in motor learning”. In: *Trends in Cognitive Sciences* 5.11, pp. 487–494.

Wolpert, Daniel M., Zoubin Ghahramani, and Michael I. Jordan (1995). “An internal model for sensorimotor integration”. In: *Science* 269.5232, pp. 1880–1882.

Wolpert, Daniel M., R Christopher Miall, and Mitsuo Kawato (1998). “Internal models in the cerebellum”. In: *Trends in Cognitive Sciences* 2.9, pp. 338–347.

Wood, Sidney (1986). “The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels”. In: *The Journal of the Acoustical Society of America* 80.2, pp. 391–401.

Wu, Wei 伍巍 (1995). “合肥话 “-i”、“-y” 音节声韵母前化探讨 [Exploration of the fronting of “-i” and “-y” rhymes in the Hefei dialect]”. In: 语言研究 *[Studies in Language and Linguistics]* 3, pp. 58–60.

Wu, Yuwen (1994). “Mandarin Segmental Phonology”. PhD thesis. University of Toronto.

Xing, Yanchun 邢晏春 (2014). 邢晏春苏州话语音词典 *[The Xing Yanchun Suzhou dialect pronunciation dictionary]*. 苏州大学出版社 [Soochow University Press].

Yan, Yiming 颜逸明 (1981). “平阳县和泰顺县的方言情况 [The dialect situation in Pingyang County and Taishun County]”. In: 方言 *[Dialect]* 1, pp. 67–72.

Yan, Yiming 颜逸明 (1988). “吴语使用人口的再统计 [New statistics on the Wú-speaking population]”. In: 中国语文 *[Zhongguo Yuwen]* 4, pp. 287–289.

Ye, Xiangling 叶祥苓 (1988). 苏州方言志 *[Record of the Suzhou Dialect]*. 江苏教育出版社 [Jiangsu Education Press].

You, Rujie (2015). “Wú 吳 Dialects”. In: *Encyclopedia of Chinese Language and Linguistics.* Ed. by Rint Sybesma. Accessed online 10 August 2017.

Yu, Alan C. L. (1999). “Aerodynamic constraints on sound change: The case of syllabic sibilants”. In: *The Journal of the Acoustical Society of America* 105.2, pp. 1096–1097.

Yuan, Jiahong and Mark Liberman (2008). “Speaker identification on the SCOTUS Corpus”. In: *Journal of the Acoustical Society of America* 123.5, p. 3878.

Zee, Eric and Wai-Sum Lee (2007). “Vowel typology in Chinese”. In: *Proceedings of ICPhS 16*.

Zhao, Rixin 赵日新 (2007). “汉语方言中的 [i] > [ɿ] [Sound change of [i] > [ɿ] in Chinese dialects] ”. In: 中国语文 *[Zhongguo Yuwen]* 1, pp. 46–54.

Zhao, Rixin 赵日新 (2008). “安徽省的汉语方言 [An Introduction of the Chinese dialects in Anhui Province]”. In: 方言 *[Dialect]* 4, pp. 361–368.

Zhaonasitu, 照那斯图 (1981). 土族语简志 *[A concise grammar of the Tu people’s language]*. 民族出版社 [Nationalities Press].

Zhengzhang, Shangfang 郑张尚芳 (1988). “关于吴语的人口 [On the Wú-speaking population]”. In: 中国语文 *[Zhongguo Yuwen]* 4, pp. 289–291.

Zhou, Dianfu 周殿福 and Zongji 吴宗济 Wu (1963). 普通话发音图谱 *[Articulatory diagrams of Standard Chinese]*. 商务印书馆 [Commercial Press].

Zhou, Liang 周良 (2000). 苏州评弹 *[Suzhou Pingtan]*. 苏州大学出版社 [Soochow University Press].

Zhou, Xinhui, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland, and Ann Choe (2008). “A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/”. In: *The Journal of the Acoustical Society of America* 123.6, pp. 4466–4481.

Zhu, Xiaonong (2006). *A grammar of Shanghai Wu.* Vol. 66. LINCOM Europa.

Zhu, Xiaonong (2015). “Phonetics”. In: *Encyclopedia of Chinese Language and Linguistics.* Ed. by Rint Sybesma. Accessed online 01 September 2017.

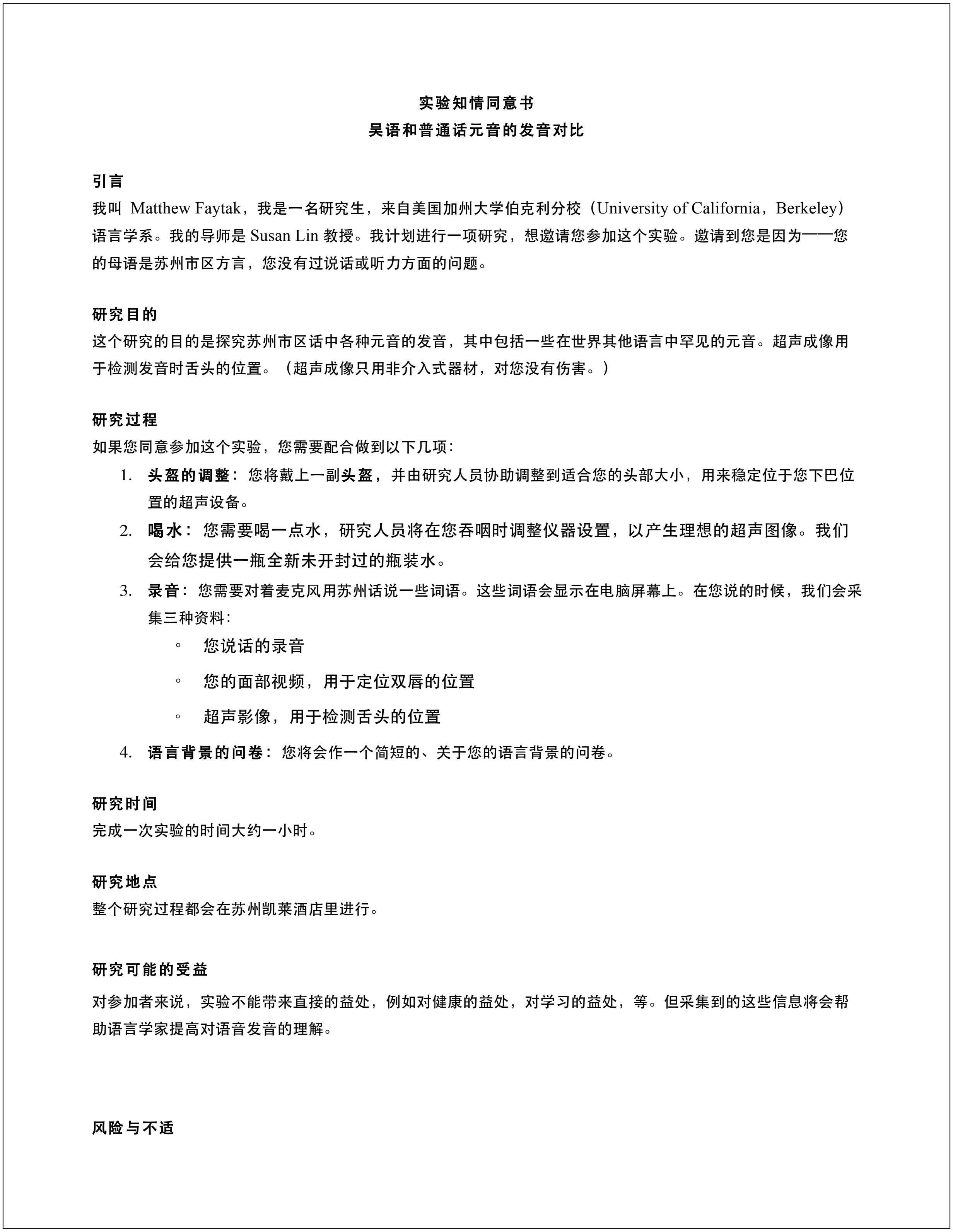
Zhu, Xiaonong 朱晓农 (2004). “汉语元音的高顶出位 [Sound changes of high vowels in Chinese dialects]”. In: 中国语文 *[Zhongguo Yuwen]* 5, pp. 440–51.

Zipf, George K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press.

Appendix A

Sūzhōu Chinese data collection instruments

Starting on the next page, the consent form, media records release, and linguistic background and demographics survey distributed to all subjects prior to the Sūzhōu Chinese experiment, as described in Chapter 4, are attached.



实验知情同意书
吴语和普通话元音的发音对比

引言

我叫 Matthew Faytak，我是一名研究生，来自美国加州大学伯克利分校（University of California，Berkeley）语言学系。我的导师是 Susan Lin 教授。我计划进行一项研究，想邀请您参加这个实验。邀请到您是因为——您的母语是苏州市区方言，您没有过说话或听力方面的问题。

研究目的

这个研究的目的是探究苏州市区话中各种元音的发音，其中包括一些在世界其他语言中罕见的元音。超声成像用于检测发音时舌头的位置。（超声成像只用非介入式器材，对您没有伤害。）

研究过程

如果您同意参加这个实验，您需要配合做到以下几项：

1. **头盔的调整**：您将戴上一副**头盔**，并由研究人员协助调整到适合您的头部大小，用来稳定位于您下巴位置的超声设备。
2. **喝水**：您需要喝一点水，研究人员将在您吞咽时调整仪器设置，以产生理想的超声图像。我们会给您提供一瓶全新未开封过的瓶装水。
3. **录音**：您需要对着麦克风用苏州话说一些词语。这些词语会显示在电脑屏幕上。在您说的时候，我们会采集三种资料：
    - 您说话的录音
    - 您的面部视频，用于定位双唇的位置
    - 超声影像，用于检测舌头的位置
4. **语言背景的问卷**：您将会作一个简短的、关于您的语言背景的问卷。

研究时间

完成一次实验的时间大约一小时。

研究地点

整个研究过程都会在苏州凯莱酒店里进行。

研究可能的受益

对参加者来说，实验不能带来直接的益处，例如对健康的益处，对学习的益处，等。但采集到的这些信息将会帮助语言学家提高对语音发音的理解。

风险与不适

与实验相关的潜在风险、副作用和不适可能包括以下这些：

1. 佩戴稳定超声设备的头盔可能导致轻微的不适。在实验过程中，如果您感觉非常不适，请先请求研究人员帮助您摘下头盔。
2. 隐私问题：与其他研究一样，我们需要有一个保密协议。我们会做好预防措施，避免出现风险（见下文）。

保密措施

为了减少泄密风险，所有您的相关数据将被存储在加州大学伯克利分校语言学系的服务器上，该服务器有密码保护。每个参加者的数据都将被加密，加密密钥只有研究人员知情。您的姓名会被替换成数字代码，姓名和数据不存在直接联系。只有研究人员了解代码和姓名之间的对应。

个人信息的使用

我们可能会在与本研究相关的出版物或公开场合提到您提供的一些个人信息（包括使用的语言、居住地、受教育程度、年龄），用于对研究的结果进行解释。在同意书的结尾，您要做出表态是否同意使用这些信息。

录音、超声图像或面部视频影像的使用：

我们也可能会在与本研究相关的出版物或公开场合提到您的录音、超声图像或以面部视频影像。 如果您同意在研究结束后使用某些资料，请阅读和签署《资料使用许可》。

研究记录的留存

我们在完成本研究后将继续保留录音，视频和超声图像，用于我们或其他研究人员进行进一步的研究。我们到实验结束后十年会留存这些研究记录。上面说明的保密措施会保护记录的安全。

补偿与费用

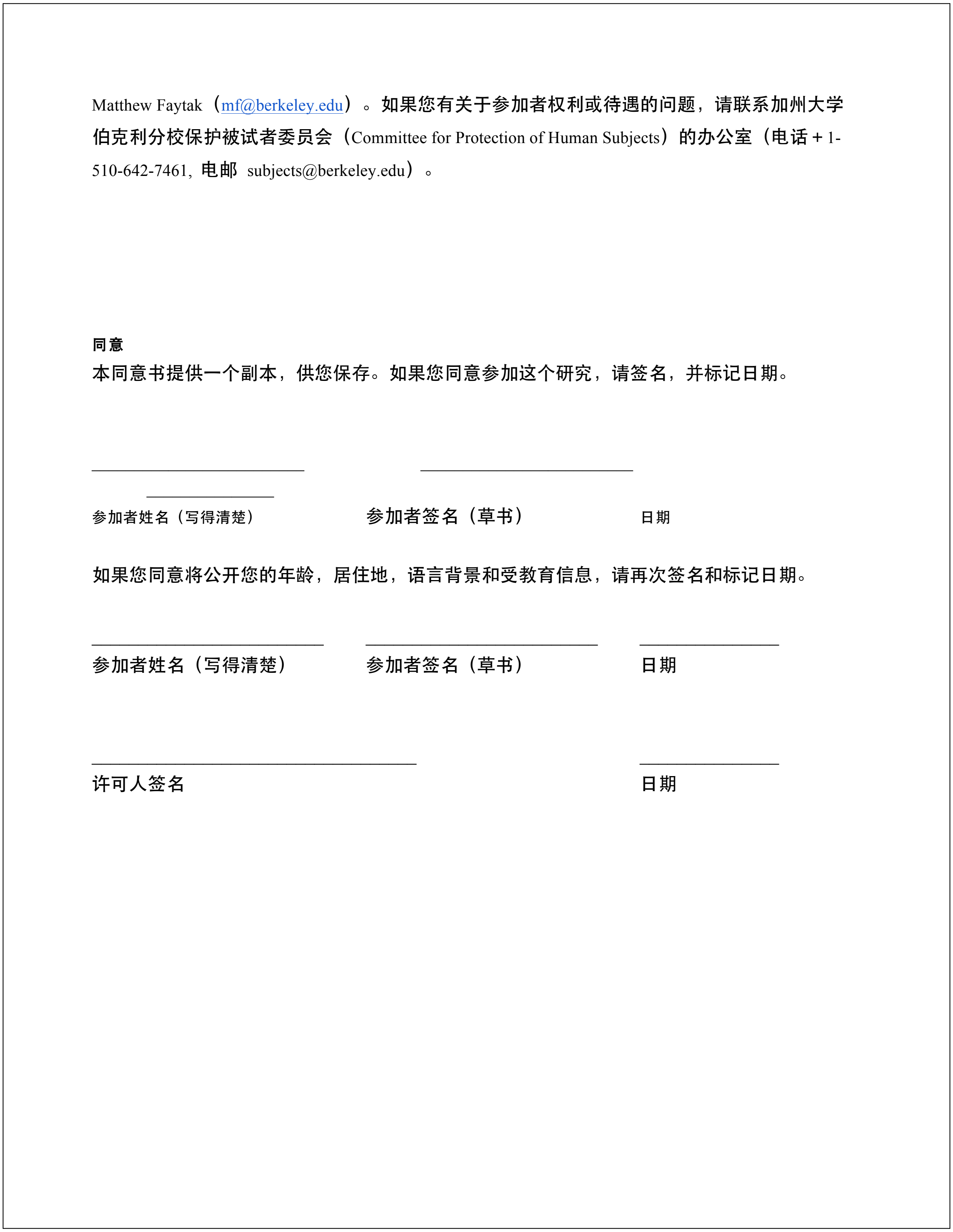
我们将支付给您劳务费 100 元。在研究结束时以现金支付。

参加者的权利

完全自愿参加。您有权拒绝参加或随时退出实验。

问题

如果您有关于本研究的任何问题，请邮件联系 Susan Lin 教授（susanlin@berkeley.edu）或学生



Matthew Faytak（mf@berkeley.edu）。如果您有关于参加者权利或待遇的问题，请联系加州大学伯克利分校保护被试者委员会（Committee for Protection of Human Subjects）的办公室（电话 + 1-510-642-7461，电邮 subjects@berkeley.edu）。

同意

本同意书提供一个副本，供您保存。如果您同意参加这个研究，请签名，并标记日期。

参加者姓名（写得清楚）　　参加者签名（草书）　　日期

如果您同意将公开您的年龄，居住地，语言背景和受教育信息，请再次签名和标记日期。

参加者姓名（写得清楚）　　参加者签名（草书）　　日期

许可人签名　　日期

资料使用许可书
吴语和普通话元音的发音对比

作为这项研究的一部分，我们将会采集您的音频、视频和超声影像。我们也会对这些数据进行分析，但我们也会将数据资料提供给其他研究者，以便他们进行观察和分析。**请在下面标出您允许其他人访问的部分。**访问度完全取决于您。我们承诺只能根据您许可的访问度使用数据。任何使用都不涉及您的姓名。

访问度由小到大排列，请标记出您对每类数据的许可：

1. 数据可以在其他实验中出现。
   录音 ________ 超声图像 ________ 视频________
   标记 标记 标记
2. 数据可以在学术出版物中出现。
   录音 ________ 超声图像 ________ 视频________
   标记 标记 标记
3. 数据可以在语言学学术会议中出现。
   录音 ________ 超声图像 ________ 视频________
   标记 标记 标记
4. 数据可以在课堂上使用。
   录音 ________ 超声图像 ________ 视频________
   标记 标记 标记
5. 数据可以在非学术的公众场合使用。
   录音 ________ 超声图像 ________ 视频________
   标记 标记 标记
6. 数据可以在电视或电台中使用。
   录音 ________ 超声图像 ________ 视频________
   标记 标记 标记

如果您同意研究人员按照以上许可度使用数据，请在下面签名，并标记日期。

________________________	________________________	______________
参加者姓名（写得清楚）	参加者签名（草书）	日期

语言背景问卷
吴语和普通话元音的发音对比

性别 (圈出)　　男　女　　　　年龄 ______岁

您的最高学历? ____________________

您曾在哪些城市或村镇居住？请先写下您的出生地，然后一列举出您连续居住一年以上的地点（城市或村镇），尽可能全面。如果写不下，您可以写在背面。

从出生　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区
从 ____岁　到____岁　在（________省）________市／镇／乡 ________区

您会说什么语言？您在什么年龄开始说这些语言？您如何评价您的阅读，写作，说话和听力理解的能力？中国方言的差异对于这项研究非常重要。请仔细列举出您能说的地方话，例如苏州话、南京话、普通话等。请不要笼统地写汉语、中文等。如果写不下，您可以写在背面。

能力水平程度的解释:
1 = 非常低，几乎不具备该项能力 4 = 一般水平　　7 = 非常熟练，当地人的水平

语言或方言	开始的年龄	能力水平（圈出):			
		阅读	写作	说话	听力理解
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
________	____岁	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

Appendix B

Sūzhōu Chinese subject metadata

In Table B.1 below, the demographic, educational, and linguistic background of subjects in the Sūzhōu Chinese study described in Chapter 4 are included in full. Several explanatory notes are required to fully appreciate the nature of the descriptors, particularly for education, given that China's secondary and postsecondary educational institutions are organized differently than in the United States. All participants have completed at least some secondary education (equivalent to high school in the United States and indicated as such). Participants marked with "UG" have completed some postsecondary education. This educational attainment takes two forms: a degree, equivalent to a bachelor's degree in the United States (indicated as "UG degree"), which is the result of a course of study lasting four to five years; and a diploma, somewhat similar to an associate's degree in the United States (indicated as "UG diploma"), which is the result of a shorter course of study of two to three years. Other postsecondary education levels (i.e., master's degree) correspond more straightforwardly to those encountered in the United States.

The location of participants' current residence is provided in the rightmost column. Most participants have resided in the city of Sūzhōu for their entire lives, and as such listing their current district of residence in the city of Sūzhōu provides sufficient information. All participants were born in one of the four urban districts of Sūzhōu city, and only a handful of participants have moved from place to place before settling back in Sūzhōu city or nearby Shànghǎi. Participants reported having resided in Shànghǎi (Speakers 1, 2, 3, 37, 44), 南京 Nánjīng (10), 广州 Guāngzhōu (18), 连云港 Lianńyúngǎng (25), 北京 Běijīng (34), and 张家港 Zhāngjiāgǎng (18), which is a sub-city under the larger political unit of Sūzhōu but outside of the Sūzhōu Chinese-speaking area proper. All five participants who have some residential history in Shànghǎi currently reside in Shànghǎi. Two participants have spent substantial time living abroad, one at a liberal arts college in the midwestern United States (34) and one in Manchester in the United Kingdom (37).

Table B.1: Participant numerical ID, gender, age, highest educational level attained, languages spoken to any degree, and location of current residence.

ID	Gender	Age	Education	Languages	Residence
1	F	20	Some college	Sūzhōu Chinese, Standard Chinese, English, German	Shanghai
2	F	22	Some college	Sūzhōu Chinese, Standard Chinese, English	Shanghai
3	F	20	Some college	Sūzhōu Chinese, Standard Chinese, English	Shanghai
4	F	48	UG diploma	Sūzhōu Chinese, Standard Chinese	Gusu
5	F	23	UG degree	Sūzhōu Chinese, Standard Chinese, English, Japanese	Gusu
6	M	20	Some college	Sūzhōu Chinese, Standard Chinese, English	Wuzhong
7	F	47	High school	Sūzhōu Chinese, Standard Chinese	Gusu
8	F	45	High school	Sūzhōu Chinese, Standard Chinese	Gusu
9	M	48	Graduate degree	Sūzhōu Chinese, Standard Chinese, English	Gusu
10	M	23	UG degree	Standard Chinese, English, Sūzhōu Chinese, Nanjing Chinese	Gusu
11	M	21	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
12	F	47	High school	Sūzhōu Chinese, Standard Chinese	Wuzhong
13	F	18	Some college	Sūzhōu Chinese, Standard Chinese, English	Gusu
14	F	54	UG diploma	Sūzhōu Chinese, Standard Chinese	Gusu
15	F	54	UG degree	Sūzhōu Chinese, Standard Chinese	Gusu
16	F	41	Graduate degree	Sūzhōu Chinese, Standard Chinese	Gusu
17	F	48	UG degree	Sūzhōu Chinese, Standard Chinese	Gusu

Continued on next page

Table B.1: Participant numerical ID, gender, age, highest educational level attained, languages spoken to any degree, and location of current residence.

ID	Gender	Age	Education	Languages	Residence
18	F	54	UG diploma	Sūzhōu Chinese, Standard Chinese	ILL
19	M	21	Some college	Sūzhōu Chinese, Standard Chinese, English	Gusu
20	F	20	Some college	Sūzhōu Chinese, Standard Chinese, English	Gusu
21	F	48	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
22	F	49	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
23	F	37	UG degree	Sūzhōu Chinese, Standard Chinese, English	Gusu
24	M	57	High school	Sūzhōu Chinese, Standard Chinese	Gusu
25	F	48	Graduate degree	Sūzhōu Chinese, Nanjing Chinese, Standard Chinese	Gusu
26	F	55	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
27	M	55	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
28	M	22	Some college	Sūzhōu Chinese, Standard Chinese, English	ILL
29	M	24	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
30	F	20	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
31	F	21	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Huqiu
32	F	43	UG diploma	Sūzhōu Chinese, Standard Chinese	Gusu
33	F	45	High school	Sūzhōu Chinese, Standard Chinese	Gusu
34	M	27	Graduate degree	Sūzhōu Chinese, Standard Chinese, Shanghainese, English	Wuzhong

Continued on next page

Table B.1: Participant numerical ID, gender, age, highest educational level attained, languages spoken to any degree, and location of current residence.

ID	Gender	Age	Education	Languages	Residence
35	M	18	Some high school	Sūzhōu Chinese, Standard Chinese, English	Huqiu
36	M	18	Some high school	Sūzhōu Chinese, Standard Chinese, English	Huqiu
37	F	22	UG degree	Sūzhōu Chinese, Standard Chinese, Shanghainese, English, Cantonese, Korean, Japanese	Shanghai
38	M	39	UG degree	Sūzhōu Chinese, Standard Chinese, English	Gusu
39	M	51	High school	Sūzhōu Chinese, Standard Chinese	Gusu
40	F	50	UG diploma	Sūzhōu Chinese, Standard Chinese, English	Gusu
41	M	20	UG diploma	Sūzhōu Chinese, Standard Chinese, Shanghainese, English, Japanese	Gusu
42	F	21	UG diploma	Sūzhōu Chinese, Standard Chinese, Shanghainese, English, Shandong Chinese, Korean, Japanese	Gusu
43	F	18	Some high school	Sūzhōu Chinese, Standard Chinese, English	Gusu
44	M	24	Some grad school	Sūzhōu Chinese, Standard Chinese, English	Shanghai

Appendix C

Sūzhōu Chinese stimulus readings

Table C.1: Readings of hànzì stimuli by subject.

Item	1	2	3	4
凹	10 æ^{44}	10 æ^{44}	13 æ^{44}	13 æ^{44}
儿	10 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}	13 ȵɪ͡ʑ^{23}	13 ȵɪ͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	13 pæ^{44}	13 pæ^{44}
边	10 pi^{44}	10 pi^{44}	13 pi^{44}	13 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	13 pɪ͡ʑ^{51}	13 pɪ͡ʑ^{51}
播	10 pə͡β^{44}	10 pə͡β^{44}	13 pə͡β^{44}	13 pə͡β^{44}
疤	10 pu^{44}	10 pu^{44}	13 pu^{44}	13 pu^{44}
C	-	-	-	-
优	11 y^{523}, 1 y͡ʑ^{44}	10 y^{523}	13 y^{523}	13 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	13 fə͡v^{44}	13 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	13 i^{44}	12 i^{44}, 1 excl, 1 disc
衣	9 ɪ͡ʑ^{44}, 1 excl	10 ɪ͡ʑ^{44}	13 ɪ͡ʑ^{44}	13 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	13 sæ^{44}	13 sæ^{44}
修	10 seɪ^{44}	10 seɪ^{44}	13 seɪ^{44}	13 seɪ^{44}
鲜	10 ɕi^{44}	10 ɕi^{44}	9 ɕi^{44}, 4 si^{44}	13 si^{44}
砂	10 su^{44}, 1 seɪ^{44}	10 su^{44}	13 su^{44}	13 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	13 sɿ^{44}	12 sɿ^{44}, 1 disc
书	10 sʮ^{44}, 1 ɕy^{44}	10 sʮ^{44}	13 sʮ^{44}	13 sʮ^{44}
蛙	10 u^{44}	9 wa^{44}, 1 u^{44}	13 u^{44}	11 u^{44}, 2 wa^{44}
哑	-	-	-	-
箫	10 ɕjɛ^{44}	7 ɕjɛ^{44}, 3 sjɛ^{44}	13 ɕjɛ^{44}	13 sjɛ^{44}

Continued on next page

Table C.1: Readings of hànzì stimuli by subject.

Item	1	2	3	4
休	7 ɕy^{44}, 2 seɪ^{44}, 2 disc	7 ɕy^{44}, 5 seɪ^{44}	12 ɕy^{44}, 1 seɪ^{44}	13 ɕy^{44}
掀	10 ɕi^{44}	10 ɕjɛ^{44}	13 ɕi^{44}	13 ɕjɛ^{44}
希	10 ɕɪ͡ʑ^{44}, 1 ɕy^{44}	10 ɕɪ͡ʑ^{44}	13 ɕɪ͡ʑ^{44}	12 ɕɪ͡ʑ^{44}, 1 excl, 1 disc
西	10 ɕɪ͡ʑ^{44}, 1 ɕi^{44}	10 ɕɪ͡ʑ^{44}	13 sɪ͡ʑ^{44}	13 ɕɪ͡ʑ^{44}
靴	10 ɕy^{44}	10 ɕy^{44}	13 ɕy^{44}	13 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}, 1 ɕɪ͡ʑ^{44}	13 ɕy͡ʑ^{44}	12 ɕy͡ʑ^{44}, 1 disc
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	13 y͡ʑ^{44}	13 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	5	6	7	8
凹	10 æ^{44}	10 æ^{44}	10 æ^{44}	10 æ^{44}
儿	10 ɲɪ͡ʑ^{23}	10 ɲɪ͡ʑ^{23}	10 ɲɪ͡ʑ^{23}	10 ɲɪ͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	10 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}
播	10 pə͡ß^{44}	10 pə͡ß^{44}	10 pə͡ß^{44}	8 pə͡ß^{44}, 2 pu^{44}
疤	9 pu^{44}, 1 pə͡ß^{44}, 1 disc	10 pu^{44}	10 pu^{44}	10 pu^{44}
C	-	-	-	-
优	10 y^{523}	10 y^{523}	9 y^{523}, 1 disc	10 y^{523}
夫	9 fə͡v^{44}, 1 excl	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	10 i^{44}
衣	9 ɪ͡ʑ^{44}, 1 i^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}
修	7 seɪ^{44}, 3 ɕy^{44}	6 seɪ^{44}, 4 ɕy^{44}	9 seɪ^{44}, 1 ɕy^{44}	10 seɪ^{44}
鲜	9 ɕi^{44}, 1 ɕɪ͡ʑ^{44}	10 ɕi^{44}	10 si^{44}	8 si^{44}, 2 ɕi^{44}
砂	10 su^{44}	10 su^{44}	10 su^{44}	10 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}, 1 ɕy͡ʑ^{44}	10 sɿ^{44}	10 sɿ^{44}
书	10 sʮ^{44}	10 sʮ^{44}	10 sʮ^{44}	10 sʮ^{44}
蛙	10 u^{44}	10 u^{44}	6 u^{44}, 3 wa^{44}, 1 disc	7 u^{44}, 3 wa^{44}
哑	-	-	-	-
箫	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 sjɛ^{44}	10 ɕjɛ^{44}

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	5	6	7	8
休	9 ɕy^{44}, 1 seɪ^{44}	7 seɪ^{44}, 3 ɕy^{44}	9 ɕy^{44}, 1 disc	9 ɕy^{44}, 1 disc
掀	10 ɕi^{44}	10 ɕi^{44}	11 ɕjɛ^{44}	9 ɕjɛ^{44}, 1 disc
希	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}
西	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}	10 si͡ʑ^{44}	10 ɕi͡ʑ^{44}
靴	9 ɕy^{44}, 1 ɕi^{44}	10 ɕy^{44}	8 ɕy^{44}, 2 disc	10 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	9	10	11	12
凹	10 æ^{44}	9 æ^{44}, 1 disc	10 æ^{44}, 1 disc	10 æ^{44}
儿	10 ȵi͡ʑ^{23}	10 ȵi͡ʑ^{23}	10 ȵi͡ʑ^{23}	10 ȵi͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	10 pi^{44}
比	10 pi͡ʑ^{51}	10 pi͡ʑ^{51}	10 pi͡ʑ^{51}	10 pi͡ʑ^{51}
播	10 pə͡β^{44}	10 pə͡β^{44}	10 pə͡β^{44}	10 pə͡β^{44}
疤	10 pu^{44}	10 pu^{44}	10 pu^{44}, 1 pæ^{44}	10 pu^{44}
C	-	-	-	-
优	10 y^{523}	9 y^{523}, 1 ȵy^{523}	10 ȵy^{523}	11 y^{523}, 1 i^{44}
夫	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	10 i^{44}
衣	10 i͡ʑ^{44}	10 i͡ʑ^{44}	10 i͡ʑ^{44}	10 i͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}
修	7 seɪ^{44}, 3 ɕy^{44}	8 seɪ^{44}, 2 ɕy^{44}	10 seɪ^{44}	10 seɪ^{44}
鲜	10 si^{44}	10 ɕi^{44}	10 ɕi^{44}	10 si^{44}
砂	10 su^{44}	10 su^{44}	10 su^{44}	10 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}
书	10 sʮ^{44}	10 sʮ^{44}	9 sʮ^{44}, 1 sɿ^{44}	10 sʮ^{44}
蛙	6 u^{44}, 4 wa^{44}	10 u^{44}	10 u^{44}	10 u^{44}, 3 wa^{44}
哑	-	-	-	-
箫	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 sjɛ^{44}
休	10 ɕy^{44}	10 ɕy^{44}	10 ɕy^{44}, 1 seɪ^{44}	10 ɕy^{44}
掀	10 ɕi^{44}	10 ɕi^{44}	10 ɕi^{44}	10 ɕi^{44}
希	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}
西	4 si͡ʑ^{44}, 6 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}	10 ɕi͡ʑ^{44}

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	9	10	11	12
靴	10 ɕy^{44}	10 ɕy^{44}	10 ɕɥəʔ^{5}	10 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	13	14	15	16
凹	10 æ^{44}	10 æ^{44}	10 æ^{44}	10 æ^{44}
儿	10 ɲɪ͡ʑ^{23}	10 aʊ^{23}	10 ɲɪ͡ʑ^{23}	10 ɲɪ͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	10 pi^{44}
比	10 pɪ͡ʑ^{51}	9 pɪ͡ʑ^{51}, 1 disc	10 pɪ͡ʑ^{51}, 1 y^{523}	10 pɪ͡ʑ^{51}
播	10 pə͡β^{44}	10 pə͡β^{44}	10 pə͡β^{44}	11 pə͡β^{44}
疤	10 pu^{44}	10 pu^{44}	9 pu^{44}, 1 pæ^{44}	9 pu^{44}, 1 disc
C	-	-	-	-
优	10 y^{523}	10 y^{523}	10 y^{523}	10 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	10 i^{44}
衣	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}
修	10 seɪ^{44}	6 ɕy^{44}, 4 seɪ^{44}	10 seɪ^{44}	9 seɪ^{44}, 2 ɕy^{44}
鲜	10 si^{44}	9 si^{44}, 1 ɕi^{44}	10 si^{44}	10 ɕi^{44}
砂	10 su^{44}	10 su^{44}	10 su^{44}	10 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}
书	10 sɥ^{44}	10 sɥ^{44}	10 sɥ^{44}	10 sɥ^{44}
蛙	10 u^{44}	11 u^{44}, 2 wa^{44}	10 u^{44}	9 u^{44}, 2 wa^{44}, 1 i^{44}, 2 disc
哑	-	-	-	-
箫	10 ɕjɛ^{44}	8 ɕjɛ^{44}, 2 sjɛ^{44}	10 sjɛ^{44}	10 ɕjɛ^{44}
休	10 ɕy^{44}	10 ɕy^{44}	10 ɕy^{44}	7 ɕy^{44}, 3 seɪ^{44}
掀	9 ɕi^{44}, 1 si^{44}	10 ɕjɛ^{44}	10 ɕjɛ^{44}	6 ɕjɛ^{44}, 5 ɕi^{44}
希	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}
西	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 sɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}
靴	10 ɕɥəʔ^{5}	10 ɕy^{44}	10 ɕy^{44}	10 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	17	18	19	20
凹	10 æ^{44}	10 æ^{44}	8 æ^{44}	10 æ^{44}, 2 disc
儿	9 ɲɪ͡ʑ^{23}, 1 disc	10 ɲɪ͡ʑ^{23}	6 aɻ^{23}, 1 ɲɪ͡ʑ^{23}	12 ni^{44}
包	10 pæ^{44}	10 pæ^{44}	8 pæ^{44}	12 pæ^{44}
边	10 pi^{44}	10 pi^{44}	8 pi^{44}	12 pi^{44}
比	9 pɪ͡ʑ^{51}, 1 disc	10 pɪ͡ʑ^{51}	8 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}, 2 pi^{44}
播	10 pə͡ß^{44}	9 pə͡ß^{44}, 1 pu^{44}, 1 disc	8 pə͡ß^{44}, 1 pu^{44}	12 pə͡ß^{44}
疤	10 pu^{44}	10 pu^{44}	8 pu^{44}	10 pu^{44}, 2 disc
C	-	-	-	-
优	10 y^{523}	10 y^{523}	8 y^{523}	12 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	8 fə͡v^{44}	12 fə͡v^{44}
烟	10 i^{44}	9 i^{44}, 1 disc	8 i^{44}	12 i^{44}
衣	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	8 ɪ͡ʑ^{44}, 1 i^{44}	12 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	8 sæ^{44}	12 sæ^{44}
修	10 seɪ^{44}	10 seɪ^{44}	9 seɪ^{44}	12 ɕy^{44}
鲜	9 si^{44}, 1 ɕi^{44}	7 si^{44}, 3 ɕi^{44}	8 ɕi^{44}	12 ɕi^{44}
砂	10 su^{44}	10 su^{44}	8 su^{44}	7 su^{44}, 5 disc
丝	10 sɿ^{44}	10 sɿ^{44}	8 sɿ^{44}	12 sɿ^{44}
书	10 sʮ^{44}	10 sʮ^{44}	8 sʮ^{44}	12 sʮ^{44}
蛙	9 u^{44}, 1 ɕi^{44}	9 u^{44}, 1 disc	7 u^{44}, 2 wa^{44}	11 u^{44}, 1 wa^{44}
哑	-	-	-	-
箫	9 sjɛ^{44}, 1 ɕjɛ^{44}	10 sjɛ^{44}	8 ɕjɛ^{44}	12 ɕjɛ^{44}
休	10 ɕy^{44}	10 ɕy^{44}	8 seɪ^{44}	12 ɕy^{44}
掀	9 ɕi^{44}, 1 ɕjɛ^{44}	9 ɕjɛ^{44}, 1 ɕi^{44}	8 ɕi^{44}	12 ɕi^{44}
希	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	8 ɕɪ͡ʑ^{44}	11 ɕɪ͡ʑ^{44}, 1 disc
西	10 sɪ͡ʑ^{44}	10 sɪ͡ʑ^{44}	6 ɕɪ͡ʑ^{44}, 1 ɕi^{44}, 1 ɕy͡ʑ^{44}	10 ɕɪ͡ʑ^{44}, 1 ɕi^{44}, 1 si^{44}
靴	10 ɕy^{44}	10 ɕy^{44}	8 ɕʮəʔ^{5}	12 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	8 ɕy͡ʑ^{44}	12 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	6 y͡ʑ^{44}, 2 ɕy͡ʑ^{44}	12 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	21	22	23	24
凹	10 æ^{44}	10 æ^{44}	10 æ^{44}	9 æ^{44}, 1 disc

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	21	22	23	24
儿	10 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}	10 aʊ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	10 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}
播	9 pə͡β^{44}, 1 disc	10 pə͡β^{44}	10 pə͡β^{44}	10 pə͡β^{44}
疤	10 pu^{44}	8 pu^{44}, 2 disc	10 pu^{44}	10 pu^{44}
C	-	-	-	-
优	10 y^{523}	9 y^{523}, 1 disc	10 y^{523}	10 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	11 i^{44}	10 i^{44}	11 i^{44}	10 i^{44}, 1 excl
衣	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}, 1 ɕɥəʔ^{5}	10 sæ^{44}	10 sæ^{44}
修	10 seɪ^{44}	10 ɕy^{44}	10 ɕy^{44}	10 seɪ^{44}
鲜	10 si^{44}	9 si^{44}, 1 disc	8 ɕi^{44}, 2 si^{44}	5 si^{44}, 5 ɕi^{44}
砂	10 su^{44}	10 su^{44}	10 su^{44}	10 su^{44}
丝	9 sɿ^{44}, 1 disc	10 sɿ^{44}	10 sɿ^{44}	9 sɿ^{44}, 1 sɪ͡ʑ^{44}
书	10 sɥ^{44}	10 sɥ^{44}	10 sɥ^{44}	10 sɥ^{44}
蛙	10 u^{44}	10 u^{44}	9 u^{44}, 1 disc	9 u^{44}, 1 disc
哑	-	-	-	-
箫	10 sjɛ^{44}	10 ɕjɛ^{44}	9 ɕjɛ^{44}, 1 sjɛ^{44}	8 sjɛ^{44}, 2 ɕjɛ^{44}
休	10 ɕy^{44}	9 ɕy^{44}, 1 ɕy͡ʑ^{44}	10 ɕy^{44}	10 ɕy^{44}
掀	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 ɕi^{44}	10 ɕjɛ^{44}
希	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}
西	10 sɪ͡ʑ^{44}	9 sɪ͡ʑ^{44}, 1 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 sɪ͡ʑ^{44}
靴	10 ɕy^{44}	10 ɕy^{44}	10 ɕy^{44}	9 ɕy^{44}, 1 disc
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	25	26	27	28
凹	13 æ^{44}	10 æ^{44}	10 æ^{44}	10 æ^{44}
儿	10 ȵɪ͡ʑ^{23}	10 aɻ^{23}	8 aɻ^{23}, 2 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}, 1 pu^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	10 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	25	26	27	28
播	10 pə͡ß^{44}	10 pə͡ß^{44}	10 pə͡ß^{44}	8 poʔ^{5}, 2 pə͡ß^{44}
疤	11 pu^{44}, 1 pə͡ß^{44}	9 pu^{44}, 1 disc	10 pu^{44}	10 pu^{44}
C	-	-	-	10 ɕɪ͡ʑ^{44}
优	10 y^{523}	10 y^{523}	10 y^{523}	10 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	10 i^{44}
衣	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}
修	8 ɕy^{44}, 2 seɪ^{44}	10 seɪ^{44}	10 seɪ^{44}	10 seɪ^{44}
鲜	10 ɕi^{44}	8 ɕi^{44}, 2 si^{44}	10 si^{44}	10 ɕi^{44}
砂	10 su^{44}	10 su^{44}	10 su^{44}	10 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}
书	10 sʮ^{44}	10 sʮ^{44}	10 sʮ^{44}	10 sʮ^{44}
蛙	10 u^{44}	10 u^{44}	10 u^{44}, 1 wa^{44}	-
哑	-	-	-	10 u^{51}
箫	10 ɕjɛ^{44}	10 ɕjɛ^{44}	9 ɕjɛ^{44}, 1 sjɛ^{44}	10 ɕjɛ^{44}
休	10 ɕy^{44}	10 ɕy^{44}	10 ɕy^{44}	9 ɕy^{44}, 1 seɪ^{44}
掀	8 ɕjɛ^{44}, 3 ɕi^{44}	10 ɕjɛ^{44}	9 ɕjɛ^{44}, 1 ɕi^{44}	10 ɕi^{44}
希	9 ɕɪ͡ʑ^{44}, 1 ɕi^{44}	9 ɕɪ͡ʑ^{44}, 1 sɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}
西	10 ɕɪ͡ʑ^{44}	8 si^{44}, 2 ɕɪ͡ʑ^{44}	10 sɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}
靴	10 ɕy^{44}	10 ɕy^{44}	10 ɕy^{44}	10 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}
迂	11 y͡ʑ^{44}	11 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	29	30	31	32
凹	10 æ^{44}	10 æ^{44}	10 æ^{44}	12 æ^{44}
儿	10 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}	10 haɻ^{23}	12 ȵɪ͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	12 pæ^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	12 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	12 pɪ͡ʑ^{51}
播	10 pə͡ß^{44}	10 pə͡ß^{44}	10 poʔ^{5}	12 pə͡ß^{44}
疤	10 pu^{44}	9 pu^{44}, 1 disc	10 pu^{44}	12 pu^{44}
C	6 sɿ^{44}, 3 si^{44}, 1 sɪ͡ʑ^{44}	-	-	-

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	29	30	31	32
优	10 y^{523}	10 y^{523}	10 y^{523}	12 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}	12 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	12 i^{44}
衣	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	12 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}	12 sæ^{44}
修	9 seɪ^{44}, 1 ɕy^{44}	9 seɪ^{44}, 1 ɕy^{44}	9 seɪ^{44}, 1 ɕy^{44}	12 seɪ^{44}
鲜	10 ɕi^{44}	9 ɕi^{44}, 1 si^{44}	10 si^{44}	12 ɕi^{44}
砂	10 su^{44}	10 su^{44}	10 su^{44}	12 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}	12 sɿ^{44}
书	10 sɥ^{44}	10 sɥ^{44}	10 sɥ^{44}	12 sɥ^{44}
蛙	-	-	-	-
哑	10 u^{51}	10 u^{51}	10 u^{51}	12 u^{51}
箫	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 sjɛ^{44}	12 ɕjɛ^{44}
休	9 ɕy^{44}, 1 seɪ^{44}	7 ɕy^{44}, 3 seɪ^{44}	10 ɕy^{44}	12 ɕy^{44}
掀	10 ɕi^{44}	10 ɕi^{44}	10 ɕi^{44}	8 ɕjɛ^{44}, 3 ɕi^{44}, 1 sjɛ^{44}
希	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	9 ɕɪ͡ʑ^{44}, 1 sɿ^{44}	12 ɕɪ͡ʑ^{44}
西	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 sɿ^{44}	12 ɕɪ͡ʑ^{44}
靴	10 ɕɥəʔ^{5}	10 ɕy^{44}	9 ɕɥəʔ^{5}, 1 ɕy^{44}	12 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	12 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	9 y͡ʑ^{44}, 1 disc	9 y^{523}, 1 y͡ʑ^{44}	12 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	33	34	35	36
凹	10 æ^{44}	10 æ^{44}	8 æ^{44}, 2 u^{51}, 1 disc	10 æ^{44}
儿	10 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}	10 ȵi^{44}	10 ȵɪ͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	10 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pi^{44}	9 pɪ͡ʑ^{51}, 1 pi^{44}
播	10 pə͡ß^{44}	10 pə͡ß^{44}	10 pu^{44}	10 pə͡ß^{44}
疤	10 pu^{44}	10 pu^{44}	10 pu^{44}	10 pu^{44}
C	-	-	-	-
优	10 y^{523}	10 y^{523}	10 y^{523}	10 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	10 i^{44}

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	33	34	35	36
衣	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	7 ɪ͡ʑ^{44}, 3 i^{44}	6 ɪ͡ʑ^{44}, 4 i^{44}
烧	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}
修	9 seɪ^{44}, 1 ɕy^{44}	10 seɪ^{44}	9 seɪ^{44}, 1 ɕy^{44}	10 seɪ^{44}
鲜	7 si^{44}, 3 ɕi^{44}	9 ɕi^{44}, 1 disc	8 si^{44}, 2 ɕɪ͡ʑ^{44}	10 ɕi^{44}
砂	10 su^{44}	9 su^{44}, 1 sæ^{44}	10 su^{44}	10 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}
书	10 sɥ^{44}	10 sɥ^{44}	9 sɥ^{44}, 1 ɕy^{44}, 1 disc	10 sɥ^{44}
蛙	-	-	-	-
哑	10 u^{51}	10 u^{51}	10 u^{51}	10 u^{51}
箫	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 ɕjɛ^{44}, 1 ɕi^{44}	7 ɕjɛ^{44}, 3 seɪ^{44}
休	10 ɕy^{44}	9 ɕy^{44}, 1 seɪ^{44}	9 ɕy^{44}, 1 seɪ^{44}	8 seɪ^{44}, 1 ɕy^{44}, 1 ɕi^{44}
掀	10 ɕjɛ^{44}	10 ɕi^{44}	5 ɕjɛ^{44}, 2 ɕi^{44}, 2 seɪ^{44}, 1 ɕɪ͡ʑ^{44}, 1 si^{44}	6 ɕjɛ^{44}, 4 ɕi^{44}
希	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕi^{44}	7 ɕi^{44}, 3 sɿ^{44}
西	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	9 ɕɪ͡ʑ^{44}, 1 sɪ͡ʑ^{44}	6 ɕi^{44}, 5 sɿ^{44}
靴	10 ɕy^{44}	10 ɕɥəʔ^{5}	10 ɕy^{44}	9 ɕɥəʔ^{5}, 1 sɥ^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy^{44}	10 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y^{523}	10 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	37	38	39	40
凹	10 æ^{44}	10 æ^{44}	9 æ^{44}	10 æ^{44}
儿	10 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}	9 aɻ^{23}	10 aɻ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}
边	10 pi^{44}	10 pi^{44}	10 pi^{44}	10 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}
播	10 pə͡ß^{44}	10 pə͡ß^{44}	11 pə͡ß^{44}	10 pə͡ß^{44}
疤	10 pu^{44}	10 pu^{44}	9 pu^{44}	10 pu^{44}
C	-	-	-	-
优	10 y^{523}	10 y^{523}	10 y^{523}	10 y^{523}
夫	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	10 i^{44}
衣	12 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}, 1 y͡ʑ^{44}	10 ɪ͡ʑ^{44}

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	37	38	39	40
烧	9 sæ^{44}, 1 su^{44}, 1 disc	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}
修	8 seɪ^{44}, 2 ɕy^{44}	9 seɪ^{44}, 1 disc	8 seɪ^{44}, 1 ɕy^{44}	10 seɪ^{44}
鲜	10 ɕi^{44}	10 si^{44}	10 si^{44}	10 si^{44}
砂	10 su^{44}	10 su^{44}	9 su^{44}	10 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	9 sɿ^{44}	10 sɿ^{44}
书	10 sɥ^{44}	10 sɥ^{44}	9 sɥ^{44}	10 sɥ^{44}
蛙	-	-	-	-
哑	10 u^{51}	10 u^{51}	10 u^{51}	10 u^{51}
箫	10 ɕjɛ^{44}	10 sjɛ^{44}	7 ɕjɛ^{44}, 2 sjɛ^{44}	10 sjɛ^{44}
休	7 ɕy^{44}, 3 seɪ^{44}	9 ɕy^{44}, 1 disc	9 ɕy^{44}, 1 seɪ^{44}	10 ɕy^{44}
掀	10 ɕi^{44}	10 ɕjɛ^{44}	5 ɕi^{44}, 1 sjɛ^{44}, 1 si^{44}, 1 ɕjɛ^{44}, 1 disc	10 ɕjɛ^{44}
希	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	8 ɕɪ͡ʑ^{44}, 1 sɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}
西	9 ɕɪ͡ʑ^{44}, 1 disc	5 sɪ͡ʑ^{44}, 5 ɕɪ͡ʑ^{44}	5 sɪ͡ʑ^{44}, 4 ɕɪ͡ʑ^{44}	9 ɕɪ͡ʑ^{44}, 1 sɪ͡ʑ^{44}
靴	10 ɕy^{44}	10 ɕy^{44}	9 ɕy^{44}	10 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	9 y͡ʑ^{44}	10 y͡ʑ^{44}

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	41	42	43	44
凹	10 æ^{44}	10 æ^{44}	10 æ^{44}	10 æ^{44}
儿	10 ȵɪ͡ʑ^{23}	10 ȵɪ͡ʑ^{23}	10 aɻ^{23}	10 ȵɪ͡ʑ^{23}
包	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}	10 pæ^{44}
边	10 pi^{44}	10 pi^{44}	9 pi^{44}, 1 disc	10 pi^{44}
比	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}	10 pɪ͡ʑ^{51}
播	10 pə͡β^{44}	9 pə͡β^{44}, 1 disc	10 pə͡β^{44}	10 pə͡β^{44}
疤	10 pu^{44}	10 pu^{44}	10 pu^{44}	10 pu^{44}
C	-	-	-	-
优	10 y^{523}	6 y^{523}, 3 ȵy^{523}, 1 disc	10 y^{523}	10 y^{523}
夫	10 fə͡v^{44}	11 fə͡v^{44}	10 fə͡v^{44}	10 fə͡v^{44}
烟	10 i^{44}	10 i^{44}	10 i^{44}	10 i^{44}
衣	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}	10 ɪ͡ʑ^{44}
烧	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}	10 sæ^{44}

Continued on next page

Table C.1: (Continued) Readings of hànzì stimuli by subject.

Item	41	42	43	44
修	10 seɪ^{44}	9 seɪ^{44}, 2 ɕy^{44}	10 seɪ^{44}, 1 ɕy^{44}	10 seɪ^{44}
鲜	10 ɕi^{44}	7 si^{44}, 3 ɕi^{44}	10 ɕi^{44}	10 si^{44}
砂	10 su^{44}	10 su^{44}	10 su^{44}	10 su^{44}
丝	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}	10 sɿ^{44}
书	10 sɥ^{44}	8 ɕy͡ʑ^{44}, 2 sɥ^{44}	10 sɥ^{44}	10 sɥ^{44}
蛙	-	-	-	-
哑	10 u^{51}	10 u^{51}	10 u^{51}	10 u^{51}
箫	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 ɕjɛ^{44}	10 sjɛ^{44}
休	5 ɕy^{44}, 5 seɪ^{44}	9 seɪ^{44}, 1 ɕy^{44}	8 ɕy^{44}, 3 seɪ^{44}	10 seɪ^{44}
掀	10 ɕi^{44}	6 ɕi^{44}, 4 si^{44}	10 ɕi^{44}	10 ɕi^{44}
希	10 ɕɪ͡ʑ^{44}	8 ɕɪ͡ʑ^{44}, 2 sɿ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}
西	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 ɕɪ͡ʑ^{44}	10 sɪ͡ʑ^{44}
靴	10 ɕɥəʔ^{5}	11 ɕy^{44}	10 ɕɥəʔ^{5}	10 ɕy^{44}
虚	10 ɕy͡ʑ^{44}	11 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}	10 ɕy͡ʑ^{44}
迂	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}	10 y͡ʑ^{44}

Table C.2: Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	1	2	3	4	5	6	7	8	9	10	11
æ^{44}	10	10	13	13	10	10	10	10	10	9	10
pæ^{44}	10	10	13	13	10	10	10	10	10	10	11
pi^{44}	10	10	13	13	10	10	10	10	10	10	10
pɪ͡ʑ^{51}	10	10	13	13	10	10	10	10	10	10	10
pə͡β^{44}	10	10	13	13	11	10	10	8	10	10	10
pu^{44}	10	10	13	13	9	10	10	12	10	10	10
y^{44}	11	10	13	13	10	10	9	10	10	9	-
fə͡v^{44}	10	10	13	13	9	10	10	10	10	10	10
i^{44}	10	10	13	12	11	10	10	10	10	10	10
ɪ͡ʑ^{44}	9	10	13	13	9	10	10	10	10	10	10
sæ^{44}	10	10	13	13	10	10	10	10	10	10	10
seɪ^{44}	13	15	14	13	8	13	9	10	7	8	11
sjɛ^{44}	-	3	-	13	-	-	10	-	-	-	-
si^{44}	-	-	4	13	-	-	10	8	10	-	-
sɪ͡ʑ^{44}	-	-	13	-	-	-	10	-	4	-	-

Continued on next page

Table C.2: Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	1	2	3	4	5	6	7	8	9	10	11
su^{44}	10	10	13	12	10	10	10	10	10	10	10
sɿ^{44}	10	10	13	12	10	10	10	10	10	10	11
sʮ^{44}	10	10	13	13	10	10	10	10	10	10	9
u$^{44/51}$	10	1	13	11	10	10	6	7	6	10	10
ɕy^{44}	19	17	25	26	21	17	18	19	23	22	10
ɕjɛ^{44}	10	17	13	13	10	10	11	19	10	10	10
ɕi^{44}	21	10	22	-	20	20	-	2	10	20	20
ɕɪ͡ʑ^{44}	20	21	13	25	21	20	10	20	16	20	20
ɕʮəʔ^{5}	-	-	-	-	-	-	-	-	-	-	10
ɕy͡ʑ^{44}	10	10	13	12	10	11	10	10	10	10	10
y͡ʑ^{44}	11	10	13	13	10	10	10	10	10	10	10

Table C.2: (Continued) Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	12	13	14	15	16	17	18	19	20	21	22
æ^{44}	10	10	10	10	10	10	10	8	10	10	10
pæ^{44}	10	10	10	11	10	10	10	8	12	10	10
pi^{44}	10	10	10	10	10	10	10	8	14	10	10
pɪ͡ʑ^{44}	10	10	9	9	10	9	10	8	10	10	10
pə͡ß^{44}	10	10	10	10	11	10	9	8	12	9	10
pu^{44}	10	10	10	9	9	10	11	9	10	10	8
y^{44}	11	10	10	11	10	10	10	8	12	10	9
fə͡v^{44}	10	10	10	10	10	10	10	8	12	10	10
i^{44}	11	10	10	10	11	10	9	9	12	11	10
ɪ͡ʑ^{44}	10	10	10	10	10	10	10	8	12	10	10
sæ^{44}	10	10	10	10	10	10	10	8	12	10	10
seɪ^{44}	10	10	4	10	12	10	10	17	-	10	-
sjɛ^{44}	10	-	2	10	-	9	10	-	-	10	10
si^{44}	10	11	9	10	-	9	6	-	1	10	9
sɪ͡ʑ^{44}	-	-	-	10	-	10	10	-	-	10	10
su^{44}	10	10	10	10	10	10	10	8	7	10	10
sɿ^{44}	10	10	10	10	10	10	10	8	12	9	10

Continued on next page

Table C.2: (Continued) Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	12	13	14	15	16	17	18	19	20	21	22
sɥ^{44}	10	10	10	10	10	10	10	8	12	10	10
u^{44}	10	10	11	10	9	9	9	7	11	10	10
ɕy^{44}	20	10	26	20	19	20	20	-	36	20	29
ɕjɛ^{44}	-	10	18	10	16	2	9	8	12	10	10
ɕi^{44}	10	9	1	-	15	11	5	17	25	-	-
ɕɪ͡ʑ^{44}	20	20	20	10	20	10	10	14	21	10	10
ɕɥəʔ^{5}	-	10	-	-	-	-	-	8	-	-	1
ɕy͡ʑ^{44}	10	10	10	10	10	10	10	11	12	10	11
y͡ʑ^{44}	10	10	10	10	10	10	10	6	12	10	10

Table C.2: (Continued) Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	23	24	25	26	27	28	29	30	31	32	33
æ^{44}	10	9	13	10	10	10	10	10	10	12	10
pæ^{44}	10	10	10	10	10	10	10	10	10	12	10
pi^{44}	10	10	10	10	10	10	10	10	10	12	10
pɪ͡ʑ^{44}	10	10	10	10	10	10	10	10	10	12	10
pə͡β^{44}	10	10	11	10	10	2	10	10	-	12	10
pu^{44}	10	10	11	9	10	11	10	9	10	12	10
y^{44}	10	10	10	10	10	10	10	10	19	12	10
fə͡v^{44}	10	10	10	10	10	10	10	10	10	12	10
i^{44}	10	10	10	10	10	10	10	10	10	12	10
ɪ͡ʑ^{44}	10	10	10	10	10	10	10	10	10	12	10
sæ^{44}	10	10	10	10	10	10	10	10	10	12	10
seɪ^{44}	-	10	2	10	10	11	10	12	9	12	9
sjɛ^{44}	1	8	-	-	1	-	-	-	10	1	-
si^{44}	2	5	-	10	10	-	3	1	10	-	7
sɪ͡ʑ^{44}	-	11	-	1	10	-	1	-	-	-	-
su^{44}	10	10	10	10	10	10	10	10	10	12	10
sɿ^{44}	10	9	10	10	10	10	16	10	21	12	10
sɥ^{44}	10	10	10	10	10	10	10	10	10	12	10
u^{44}	9	9	10	10	10	10	10	10	10	12	10

Continued on next page

Table C.2: (Continued) Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	23	24	25	26	27	28	29	30	31	32	33
ɕy^{44}	30	19	28	20	20	19	10	18	12	24	21
ɕjɛ^{44}	9	12	18	20	18	10	10	10	-	20	20
ɕi^{44}	18	5	14	8	1	20	20	19	10	15	3
ɕɪ͡ʑ^{44}	20	10	19	11	10	30	20	20	9	24	20
ɕɥəʔ^{5}	-	-	-	-	-	-	10	-	9	-	-
ɕy͡ʑ^{44}	10	10	10	10	10	10	10	10	10	12	10
y͡ʑ^{44}	10	10	11	10	10	10	10	9	1	12	10

Table C.2: (Continued) Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	34	35	36	37	38	39	40	41	42	43	44
æ^{44}	10	8	10	10	10	9	10	10	10	10	10
pæ^{44}	10	10	10	10	10	10	10	10	10	10	10
pi^{44}	10	20	11	10	10	10	10	10	10	9	10
pɪ͡ʑ^{44}	10	-	9	10	10	10	10	10	10	10	10
pə͡β^{44}	10	-	10	10	10	11	10	10	9	10	10
pu^{44}	10	20	10	10	10	9	10	10	10	10	10
y^{44}	10	20	10	10	10	10	10	10	6	10	10
fə͡v^{44}	10	10	10	10	10	10	10	10	11	10	10
i^{44}	10	13	14	10	10	10	10	10	10	10	10
ɪ͡ʑ^{44}	10	7	6	12	10	10	10	10	10	10	10
sæ^{44}	11	10	10	9	10	10	10	10	10	10	10
seɪ^{44}	11	12	21	11	9	9	10	15	18	13	20
sjɛ^{44}	-	-	-	-	10	3	10	-	-	-	10
si^{44}	-	9	-	-	10	11	10	-	11	-	10
sɪ͡ʑ^{44}	-	1	-	-	5	6	1	-	-	-	10
su^{44}	9	10	10	11	10	9	10	10	10	10	10
sɿ^{44}	10	10	18	10	10	9	10	10	12	10	10
sɥ^{44}	10	9	11	10	10	9	10	10	2	10	10
u^{44}	10	12	10	10	10	10	10	10	10	10	10
ɕy^{44}	9	31	1	19	19	19	20	5	14	9	10
ɕjɛ^{44}	10	15	13	10	10	8	10	10	10	10	-

Continued on next page

Table C.2: (Continued) Total readings collected per subject, regardless of the stimulus item that resulted in the reading. Excludes readings that lack target segments and those that begin with a nasal stop.

reading	34	35	36	37	38	39	40	41	42	43	44
ɕi^{44}	19	13	28	20	-	5	-	20	9	20	10
ɕi͡ʑ^{44}	20	12	-	19	15	12	19	20	18	20	10
ɕʮəʔ^{5}	10	-	9	-	-	-	-	10	-	10	-
ɕy͡ʑ^{44}	10	-	10	10	10	10	10	10	19	10	10
y͡ʑ^{44}	10	-	10	10	10	10	10	10	10	10	10

Table C.3: Tokens of target segments collected by subject.

Target	1	2	3	4	5	6	7	8	9	10	11
/s/	53	58	83	89	48	53	79	58	61	48	51
/ɕ/	80	75	86	76	82	78	49	70	69	82	80
/i/	41	30	52	38	41	40	30	30	40	40	40
/y/	30	27	38	39	31	27	27	29	33	31	10
/i͡ʑ/	39	41	52	51	40	40	40	40	40	40	40
/y͡ʑ/	21	20	26	25	20	21	20	20	20	20	20
/ɿ/	10	10	13	12	10	10	10	10	10	10	11
[ʮ]	10	10	13	13	10	10	10	10	10	10	9
/æ/	30	30	39	39	30	30	30	30	30	29	31
/u/	30	21	39	35	29	30	26	29	26	30	30
[ə͡v]	10	10	13	13	9	10	10	10	10	10	10
[ə͡ß]	10	10	13	13	11	10	10	8	10	10	10

Table C.3: (Continued) Tokens of target segments collected by subject.

Target	12	13	14	15	16	17	18	19	20	21	22
/s/	70	61	55	80	52	78	76	49	44	79	69
/ɕ/	60	69	75	50	80	53	54	58	106	50	61
/i/	41	40	30	30	36	40	30	34	52	31	29
/y/	31	20	36	31	29	30	30	8	48	30	38
/i͡ʑ/	40	40	39	39	40	39	40	30	43	40	40
/y͡ʑ/	20	20	20	20	20	20	20	17	24	20	21
/ɿ/	10	10	10	10	10	10	10	8	12	9	10

Continued on next page

Table C.3: (Continued) Tokens of target segments collected by subject.

Target	12	13	14	15	16	17	18	19	20	21	22
[ɥ]	10	10	10	10	10	10	10	8	12	10	10
/æ/	30	30	30	31	30	30	30	24	34	30	30
/u/	30	30	31	29	28	29	30	24	28	30	28
[ə͡v]	10	10	10	10	10	10	10	8	12	10	10
[ə͡ß]	10	10	10	10	11	10	9	8	12	9	10

Table C.3: (Continued) Tokens of target segments collected by subject.

Target	23	24	25	26	27	28	29	30	31	32	33
/s/	43	73	42	61	71	51	60	53	80	61	56
/ɕ/	87	56	89	69	59	89	80	77	50	95	74
/i/	40	30	34	38	31	40	43	40	40	39	30
/y/	40	29	38	30	30	29	20	28	31	36	31
/i͡ʑ/	40	41	39	32	40	50	41	40	29	48	40
/y͡ʑ/	20	20	21	20	20	20	20	19	11	24	20
/ɿ/	10	9	10	10	10	10	16	10	21	12	10
[ɥ]	10	10	10	10	10	10	10	10	10	12	10
/æ/	30	29	33	30	30	30	30	30	30	36	30
/u/	29	29	31	29	30	31	30	29	30	36	30
[ə͡v]	10	10	10	10	10	10	10	10	10	12	10
[ə͡ß]	10	10	11	10	10	2	10	10	-	12	10

Table C.3: (Continued) Tokens of target segments collected by subject.

Target	34	35	36	37	38	39	40	41	42	43	44
/s/	51	61	70	51	74	66	71	55	63	53	90
/ɕ/	78	71	61	78	54	54	59	75	70	79	40
/i/	39	55	53	40	30	36	30	40	40	39	40
/y/	19	51	11	29	29	29	30	15	20	19	20
/i͡ʑ/	40	20	15	41	40	38	40	40	38	40	40
/y͡ʑ/	20	-	20	20	20	20	20	20	29	20	20
/ɿ/	10	10	18	10	10	9	10	10	12	10	10
[ɥ]	10	9	11	10	10	9	10	10	2	10	10
/æ/	31	28	30	29	30	29	30	30	30	30	30

Continued on next page

Table C.3: (Continued) Tokens of target segments collected by subject.

Target	34	35	36	37	38	39	40	41	42	43	44
/u/	29	42	30	31	30	28	30	30	30	30	30
[ə͡v]	10	10	10	10	10	10	10	10	11	10	10
[ə͡β]	10	-	10	10	10	11	10	10	9	10	10