THE UNIVERSITY OF CHICAGO

SIBILANT CATEGORIZATION, CONVERGENCE, AND CHANGE: THE CASE OF /S/-RETRACTION IN AMERICAN ENGLISH

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF LINGUISTICS

BY
JACOB BRENNAN PHILLIPS

CHICAGO, ILLINOIS
JUNE 2020

I'd rather learn from one bird how to sing

than teach ten thousand stars how not to dance

– E.E. Cummings

TABLE OF CONTENTS

LIST OF FIGURES vii

LIST OF TABLES ix

ACKNOWLEDGMENTS x

ABSTRACT xiv

1 INTRODUCTION 1
- 1.1 Definition of key terms 3
- 1.2 Notes on phonetic transcriptions 4
- 1.3 Outline of chapters 5

2 /S/-RETRACTION: PERCEPTION, PRODUCTION, & ORIGINS 7
- 2.1 Production of /s/-retraction 7
  - 2.1.1 Phonological distribution of /s/-retraction 8
  - 2.1.2 Acoustics 11
  - 2.1.3 Articulation 14
- 2.2 Perception of /s/-retraction 17
- 2.3 Theories regarding the origins of /s/-retraction 19
  - 2.3.1 Coarticulation 20
  - 2.3.2 Assimilation at a distance 23
  - 2.3.3 Local assimilation 24
- 2.4 Theories regarding the current grammatical status of /s/-retraction 30
- 2.5 Sociolinguistic distribution of /s/-retraction 32
  - 2.5.1 Apparent time studies 33
  - 2.5.2 Geographic distribution 34
  - 2.5.3 Gender distribution 35
- 2.6 Potential socio-indexical meaning of /s/-retraction 36
  - 2.6.1 Prevocalic or context-independent socio-indexicality of sibilants . . . 37
  - 2.6.2 Preconsonantal socio-indexicality of sibilants 38
  - 2.6.3 Online meta-commentary surrounding /s/-retraction 41

3 EXPERIMENT I: CUE INTEGRATION 48
- 3.1 An introduction to cue integration 49
- 3.2 Study overview 52
- 3.3 Methods & materials 54
  - 3.3.1 Stimuli 54
  - 3.3.2 Participants & procedure 61
  - 3.3.3 Measurements, analyses, & hypotheses 68
- 3.4 Results 74
  - 3.4.1 Prevocalic results 75

3.4.2 Preconsonantal results . . . 79
3.5 Discussion . . . 91

4 EXPERIMENT II: CATEGORIZATION . . . 98
4.1 An introduction to phoneme categorization . . . 99
4.1.1 Compensation for coarticulation . . . 99
4.1.2 Adjustments for perceived speaker attributes . . . 102
4.2 Study overview . . . 103
4.3 Methods & materials . . . 104
4.3.1 Stimuli materials . . . 104
4.3.2 Participants & procedure . . . 107
4.3.3 Hypotheses & analysis . . . 110
4.4 Results . . . 113
4.5 Discussion . . . 121

5 EXPERIMENT III: CONVERGENCE . . . 126
5.1 An introduction to convergence . . . 127
5.1.1 Convergence outside the laboratory . . . 130
5.1.2 Convergence in the laboratory . . . 134
5.1.3 Convergence and exemplar theory . . . 139
5.1.4 Convergence and Communication Accommodation Theory . . . 140
5.2 Study overview . . . 141
5.3 Methods & materials . . . 142
5.3.1 Stimuli . . . 142
5.3.2 Participants & procedure . . . 146
5.3.3 Post-processing . . . 150
5.3.4 Measurements, analyses, & hypotheses . . . 151
5.4 Results . . . 157
5.4.1 Retraction Ratio results . . . 158
5.4.2 Difference in Distance results . . . 165
5.4.3 Direction of Shift results . . . 170
5.4.4 Individual results . . . 174
5.5 Discussion . . . 181

6 CONCLUSIONS & IMPLICATIONS . . . 189
6.1 Summary of reported experiments . . . 189
6.2 General discussion & conclusions . . . 192
6.3 Limitations & future directions . . . 196

A APPENDIX . . . 198
A.1 Complete statistical models . . . 198
A.2 Stimuli materials . . . 205
A.3 Online meta-discourse sources . . . 209
A.4 Cognitive, personality, & demographic surveys . . . 220

A.4.1 Basic demographic questions 220
A.4.2 Neurological, language, and hearing questions 221
A.4.3 Big Five 221
A.4.4 Empathy Quotient 223
A.4.5 MRAS 224
A.4.6 Promis Anxiety 225
A.4.7 Experimental Impressions 226

REFERENCES 227

LIST OF FIGURES

2.1 Centroid frequency for different sibilant onsets 9
2.2 Example spectral slices by phonological environment 12
2.3 Retraction ratio values for different sibilant onsets 14
2.4 Predicted sexuality from social evaluation of /sCr/ cluster (Phillips, 2018) . . . 41
2.5 Example images of online meta-commentary 44
2.6 Indexical field for /s/-retraction from online meta-commentary 46

3.1 Sample visual stimuli for eye-tracking 60
3.2 Geographic distribution of Experiment I participants 61
3.3 Sample trial for Experiment I 64
3.4 Item accuracy for Experiment I 70
3.5 Fixation proportion for prevocalic sibilants 78
3.6 Fixation proportion by retraction condition 82
3.7 Fixation proportion by place of articulation 84
3.8 Fixation proportion by cluster and retraction condition 85
3.9 Individual variation in fixation in the decreased retraction condition 87
3.10 Individual variation in fixation in the increased retraction condition 88
3.11 Individual variation in fixation in the hyper-increased retraction condition . . . 89

4.1 Toy example of perceptual compensation 100
4.2 Sample faces for categorization task 106
4.3 Geographic distribution of Experiment II participants 108
4.4 Sample trial for Experiment II 109
4.5 Random intercepts and trial slopes in categorization 114
4.6 Predicted categorization by step and cluster 115
4.7 Predicted categorization by step and age 116
4.8 Predicted categorization by step and indicators of toughness 117
4.9 Predicted categorization by step, cluster and toughness endorsement 119
4.10 Predicted categorization by step, age and toughness endorsement 120

5.1 Geographic distribution of Experiment III participants 147
5.2 Pre-test retraction ratios for the different clusters 158
5.3 Pre-test retraction ratios for retractors and non-retractors 159
5.4 Pre-test retraction ratios relative to the model talker 160
5.5 Pre-test and post-test retraction ratios 163
5.6 Pre-test and post-test retraction ratios by baseline relation 164
5.7 Difference in distance by baseline relation 168
5.8 Direction of shift or each cluster target 172
5.9 Direction of shift by baseline relation 173
5.10 Pre-test and post-test retraction ratios 175
5.11 Individual /str/ retraction ratios in the increased retraction condition 177
5.12 Individual /skr/ retraction ratios in the increased retraction condition 178

5.13 Individual /spr/ retraction ratios in the increased retraction condition 179
5.14 Individual /str/ retraction ratios in the decreased retraction condition 180

LIST OF TABLES

2.1 Common meta-discourse associations 43

3.1 Sibilant-initial wordlist for Experiment I: Cue Integration 55
3.2 Model talker values by phonological environment 56
3.3 Scaling factors used in sibilant stimuli creation 57
3.4 Stop-initial wordlist for Experiment I: Cue Integration 58
3.5 Mean naturally-produced voice onset time values for the model talker 59
3.6 Voice onset time for increased and decreased VOT conditions 59
3.7 Pairing of sibilant- and stop initial stimuli for Experiment I 63
3.8 Experiment I: Cue Integration model predictions for /s/ vs. /ʃ/ fixations 76
3.9 Experiment I: Cue Integration model predictions for /sC/ vs. /sCr/ fixations . . 81

4.1 Faces and their normed ratings 107
4.2 Experiment II: Categorization model predictions 113

5.1 Sibilant-initial wordlist for Experiment III: Convergence 143
5.2 Stop-initial wordlist for Experiment III: Convergence 145
5.3 Experiment III: Convergence model predictions for retraction ratio 162
5.4 Experiment III: Convergence model predictions for difference in distance 167
5.5 Experiment III: Convergence model predictions for direction of shift 171

A.1 Experiment I: Complete prevocalic cue integration model predictions 198
A.2 Experiment I: Complete preconsonantal cue integration model predictions 199
A.3 Experiment II: Complete categorization model predictions 200
A.4 Experiment III: Complete retraction ratio predictions 201
A.5 Experiment III: Complete difference in distance predictions 203
A.6 Experiment III: Complete direction of shift predictions 204
A.7 Complete list of stimuli for Experiment I and III 205

ACKNOWLEDGMENTS

While this dissertation was finished in a period of self-isolation and social distancing, the entire journey would not have been possible without the very real support of my many mentors, teachers, colleagues, friends, and family.

I've been incredibly fortunate to work with a committee who constantly challenges and supports me. As a first year graduate student with no idea how to design an experiment to answer the questions I was only beginning to ask, Alan Yu welcomed me into his lab and involved me in his research projects. Since then, as my advisor and committee chair, Alan has given me the guidance and support to develop my own research agenda. Thank you for your inexhaustible wisdom and feedback throughout this journey. Diane Brentari, who many years ago introduced me to the fundamental concepts in phonology that are the bedrock to this dissertation, thank you for challenging me to approach my findings from many different perspectives. Susan Lin, who welcomed me to her lab at UC-Berkeley during my fourth year, thank you not only the technical support you gave me to learn new methodologies, but also the emotional support to work through their inevitable challenges and frustrations. Jane Stuart-Smith, who joined this project from across the pond, thank you for expressing an interest in my work and lifting up new career researchers like me who look to you as a role model.

I am also thankful to have the support of so many other people at the University of Chicago. Lenore Grenoble and Peggy Mason, thank you for giving me the opportunity of a lifetime to work on a project that challenges me in a very different way than my dissertation research. Thank you for putting your confidence in me and treating me as a collaborator. And Kim, thank you for putting up with my repetitive experiments and questions with an inquisitive mind and a good attitude. Ming Xiang, thank you for giving me guidance as I ventured into speech perception research and for allowing me to use your equipment to ask the questions that most excite me. Laura Staum Casasanto, thank you for guiding me

in my journey into sociolinguistic research and helping me understand the questions worth asking. Karlos Arregi and Greg Kobele, thank you for advising my first qualifying paper as I briefly ventured into morphosyntax. And to so many others in the Linguistics Department at UChicago, thank you for our conversations and the lessons you've taught me, both inside and outside the classroom.

My success in graduate school is also due in large part to the supportive environment at Swarthmore College. Donna Jo Napoli, who taught the first linguistics class I ever took, thank you for the unwavering support through my undergraduate career. Nathan Sanders, thank you for introducing me to phonetics and providing me with a role model for what a devoted teacher looks like. And K. David Harrison, thank you for guiding me through my first extended research project and for being the first professor to treat me as a collaborator. Without all of your support, the journey would have been over before it even started.

My research is much stronger today because of the support and conversations of the broader linguistics community. I extend my gratitude particularly to Sarah Bakst, Patrice Beddor, Kathryn Campbell-Kibler, Eleanor Chodroff, Andries Coetzee, Annette D'Onofrio, Matthew Faytak, Sharon Inkelas, Keith Johnson, Bob McMurray, Lyra Magloughlin, Jeff Mielke, Anne Pycha, Timo Roettger, Meredith Tamminga, Eric Wilbanks, and Georgia Zellou. I would also like to thank the anonymous reviewers for the National Science Foundation who both believed in my research and also helped me build a stronger research agenda through their critiques. On that note, I want to thank the National Science Foundation, the University of Chicago and the Mellon Foundation for providing the financial resources to make my dissertation research possible. The research in this dissertation would not have been possible without the help of my dedicated research assistants over the years. Thank you Max Fennell-Chametzky, Sasha Elenko, Giovanna Hooton, Josef Klafka, Paige Resnick, Thomas Sostarics, and Hillel Steinmetz for your hard work and collaboration. Additional thanks to Paige Resnick who caught the hundreds of typos that filled this dissertation and

helped cut down some of the obscenely long sentences that I have a knack for writing.

I would like to thank my fellow graduate students who challenged me, supported me, and laughed with and/or at me: Carissa Abrego-Collier, Ksenia Ershova, Josh Falk, Stephanie Locke, Aurora Martinez, Kat Montemurro, Asia Pietraszko, Betsy Pillion, and Adam Singerman. In particular, graduate school would have been insurmountable without three of my fellow graduate students. Katie Franich, you were a second advisor to me. Thank you for your tea dates, tireless advice, and moral support. Jeff Geiger, thank you for lunch club, gossip club, complaining club, and every other club. Emily Hanink, thanking for being my plus one to the ballet, homemade soap provider, recipe taste tester, and emergency facetime contact no matter what time difference is between us.

Up to this point I've only thanked linguists, but my dissertation research would not have been possible without the many non-linguists who give my life balance. To my muggiest of friends: Nitya Kadambi, Revathi Kollipara, and Sirisha Tummala, thanking for laughing at my jokes when no one else will. Henry Blood, Michael Frasco, and Gareth Jones, thank you for pushing me in both training and intense, and at time combative, discussion. Rohita Kadambi and Janie Bryant, thank you for being as excited about musical theater as I am. Allison Ranshous, thank you for always making me feel welcome in New York and being unafraid to look a fool alongside me. Taryn Colonnese, thank you for seven solid years of bookclub, for long meandering walks, and for planning our house with two gardens. Rebecca Hammond, thank you for our adventures and for always being excited by whatever new thing excites me, just as I will always be excited to do the same for you.

Joshua Prenner – okay Yusha, it felt weird even writing that – deserves particular acknowledgment for getting me through some of the hardest parts of this dissertation. Thank you for always believing in me even when I didn't believe in myself. You fry up tofu better than I ever could and put up with watching the British murder mysteries that have gotten me through my bouts of writer's (and experimenter's) block. Even if I sometimes disparagingly

remark that there's no comparison between medical school rotations and dissertating, your dedication and determination in your own career has been a constant source of inspiration to me throughout this process. And Joanie, thank you for snuggles and distractions. You made the final year of writing this dissertation so much more enjoyable and survivable!

And finally, I'm eternally grateful to my family, who has supported me every step of the way. Thank you for your love, kindness, and understanding. Mom, thank you for your unwavering, endless love and for instilling in me a love of home cooked food, good television, and the feeling of getting my hands dirty in the garden. I know I said I'd never apologize for having friends over for pizza and a movie when you were preparing for your dissertation defense, but maybe I've changed my mind now. Gary, thank you for finding the best restaurants and haberdashery adventures around Chicago, which were always the perfect cure to a rough week in the lab. Dad, thank you for teaching me to always count hawks on long drives and how to identify wildflowers on spring hikes through the woods. I couldn't have made it through grad school without escaping the city every once in a while. Penney, I'm so happy that you share my happy place and have provided me a writing retreat complete with a dissertation desk, even if the floor in front of the fire got more use. To my siblings and siblings-in-law: Jordan, Marty, Taylor, Nicole, Ben, and Jason, thank you for so many adventures and experiences over the past thirty years, from tobogganing down the stairs to making home videos, from hiking up mountains to jumping off cliffs, from picking apples every September to baking Nana's pies. While there may have been some bumps along the road, like the time you tied me up and locked me in the bathroom (for being unbearably annoying, I'm sure), I'm so lucky to have had so many role models to turn to for guidance, support, and friendship. And lastly, to my nieces and nephews: Caroline, Jack, Harrison, and Sloane, thank you for being a boundless source of energy and love. I can't wait to see the princes and princesses, astronauts and artists, paleontologists and herpetologists you'll turn into!

ABSTRACT

This dissertation examines /s/-retraction, a sound change in progress by which /s/ approaches /ʃ/ in the context of /r/, such that *street* may sound more like *shtreet*. Previous research on this phenomenon has focused largely on the production of /str/ clusters, asking how /str/ is articulated, how it is realized acoustically, and what phonological process ultimately leads to a more /ʃ/-like /s/. This dissertation contributes to the understanding of the phenomenon by turning to the perception of /s/-retraction. Specifically, this dissertation seeks to answer two fundamental questions: Do listeners have detailed phonological knowledge about /s/-retraction? And if so, how do they use that knowledge? The results of three different experiments are presented that examine different aspects of how listeners may use their knowledge about /s/-retraction and what that might tell us about the origins, grammatical status, and trajectory of /s/-retraction as a sound change and the transition and propagation of sound change in general. Firstly, Experiment I asks how that knowledge influences speech processing. Using a lexical identification task with eye tracking, Experiment I finds that listeners can use the cues of /s/-retraction in order to anticipate the upcoming /r/. Secondly, Experiment II asks how that knowledge influences sociolinguistic perception. Using a phoneme categorization task, Experiment II finds that only listeners who most strongly endorse traditional stereotypes of masculinity are likely to attribute a retracted /s/ to a performance of masculine toughness. Finally, Experiment III asks how the perception of /s/-retraction influences an individual's own production when they take a turn as a speaker. Using a covert shadowing task, Experiment III finds that listeners converge toward manipulated degrees of /s/-retraction, but only if the model talker exhibits a pattern sufficiently similar to their own. Taken together, the findings for /s/-retraction challenge a convergence path to sound change, by which conversational shifts persist and accumulate to lead to lasting change, but support a coarticulatory path to sound change, by which listeners gradually shift the cues of coarticulation from the source to the target.

CHAPTER 1

INTRODUCTION

Variation is a natural and ubiquitous feature of language, and of the acoustic signal in particular. Each utterance is distinct from the last and each speaker is distinct from their interlocutor. This variation is both large and small, predictable and unpredictable. Nonetheless, listeners are able to consistently and effortlessly parse this ever-varying speech stream into discrete and meaningful words and sounds.

One such source of variation is *coarticulation*, which describes a scenario where a speech sound is influenced by, and thus becomes more similar to, a neighboring speech sound. Over time, the relative degree of coarticulation can increase and ultimately lead to categorical sound change. This dissertation is concerned with the perception and production of speech in the intermediate stages between stable coarticulation and categorical sound change. In particular, this dissertation asks how listeners account for the coarticulatory variation in their perception and how that may influence their own production when they respond in turn. With these questions, this work seeks to shed light on the *transition problem* (Weinreich et al., 1968), which asks how a sound change proceeds from one stage to another over time. Tied up in the transition problem is how the sound change propagates, or how it spreads from speaker to speaker and community to community.

To address these questions, I examine /s/-retraction, a sound change in progress in American English by which /s/ approaches /ʃ/ in the context of /r/, especially in /str/ clusters. Thus, for a speaker exhibiting /s/-retraction, *street* may sound more like *shtreet*. The nature of /s/-retraction is explored in more detail later in this dissertation, but at its core /s/-retraction can be thought of a process by which /s/ is influenced by the upcoming /r/ in such a way that the tongue body retracts and the lips protrude. These small articulatory changes can have outsized acoustic consequences, resulting in an onset sibilant that may be heard as /ʃ/. Of particular note, /s/-retraction is not observed to the same degree or with

the same frequency in /spr/ or /skr/ clusters, such that *scream* rarely sounds like *shcream*.

Much of the previous attention that /s/-retraction as a phenomenon has received has been focused on the production of /str/ clusters. This dissertation contributes to the understanding of the sound change by giving particular attention to the perception of /s/-retraction. Specifically, this dissertation asks the fundamental question, do listeners have detailed phonological knowledge about this process? And if they do, do listeners use that detailed phonological knowledge? This dissertation employs three different experiments that each seeks to examine a different aspect of how listeners can use their detailed phonological knowledge about the phenomenon and what that might tell us about the origins, grammatical status, and trajectory of /s/-retraction as a sound change and the transition and propagation of sound change in general.

- Firstly, this dissertation asks if listeners use /s/-retraction in speech processing. Specifically, using eye tracking during a lexical identification task, Experiment I tests whether listeners can use the cues of retraction in order to accurately predict the presence of an upcoming /r/. This experiment highlights the cue weight that listeners assign to /s/-retraction and how that might shed light on the trajectory of the sound change.

- Secondly, this dissertation asks if listeners use /s/-retraction as a social marker. Specifically, using a categorization task of ambiguous stimuli presented with varying indicators of masculinity, Experiment II tests whether listeners attribute acoustic retraction to performances of masculinity and toughness. This experiment highlights the potential socio-indexical role of /s/-retraction and how that might contribute to the propagation of the sound change.

- Finally, this dissertation asks if listeners adjust their own production as a result to the /s/-retraction they perceive in an interlocutor. Using a covert shadowing task, Experiment III tests whether a participant will converge toward a model talker with a

manipulated degree of retraction. This experiment assesses how the short-term shifts we produce during a conversation may ultimately be a path to lasting, categorical sound change.

In the remainder of this chapter, I provide a definition of key terms to this work (Section 1.1), notes on the use of phonetic transcriptions (Section 1.2), and a road map to the structure of this dissertation (Section 1.3).

1.1 Definition of key terms

This dissertation is concerned with *sibilant* speech sounds, which are a class of fricative consonants characterized by their high intensity and high frequency relative to other speech sounds. In particular, I will focus on the voiceless alveolar fricative /s/, as in ***s****ip* and ***s****uit*, and the voiceless palato-alveolar fricative /ʃ/, as in ***sh****ip* and ***sh****oot*.

Furthermore, this dissertation is concerned with *retraction*. In this dissertation, I will use the terms *retraction* and *retracted* to refer to both the articulatory act of pulling the tongue back in the oral cavity as well as its acoustic consequences. Specifically, I will refer to a sibilant as being *retracted* when it is naturally produced or artificially manipulated in such a way as to have a lower centroid frequency.

In this dissertation, I will use the term *convergence* (Natale, 1975a) to refer to the process by which a speaker adopts some of the characteristics of their interlocutor. This process is also commonly referred to as *imitation* (Goldinger, 1998), *accommodation* (Giles et al., 1991), and less commonly *adaptation* (Gregory & Hoyt, 1982), *alignment* (Garrod & Pickering, 2004), *entrainment* (Brennan & Clark, 1996), *coordination* (Garrod & Anderson, 1987), and *persistence* (Bock, 1986). *Alignment*, *entrainment*, *coordination*, and *persistence* typically refer to syntactic, semantic, or lexical shifts, while *convergence*, *imitation*, and *accommodation* typically refer to phonetic shifts. Although these latter three terms are frequently used interchangeably, they may be teased apart in the degree to which they suggest consciousness

or directionality. *Imitation* is often used to describe a voluntary action, *accommodation* often describes conscious and subjective shifts, and both terms suggest a unidirectional shift exhibited by one interlocutor toward another. *Convergence*, however, is largely agnostic to whether the shift was voluntary, conscious, or subjective, and whether one or both interlocutors exhibited shifts; it is simply concerned with whether any shifts in the phonetic signal can be observed (Sonderegger, 2012). As this dissertation contains a shadowing task that is covert by design, I have selected the term *convergence* to be agnostic as to whether the participants are intentionally or unintentionally shifting their speech toward the model talker.

1.2 Notes on phonetic transcriptions

Throughout this dissertation, I will use the International Phonetic Alphabet with a few key exceptions. Firstly, I will use ⟨r⟩ as a catch-all symbol to represent the American English rhotic characterized orthographically as "r", regardless of whether the articulation of this sound is either bunched or retroflex (Delattre & Freeman, 1968; Mielke et al., 2010). When necessary to distinguish between these different articulations of /r/, I will use ⟨ɹ⟩ to represent the bunched alveolar approximant and ⟨ɻ⟩ to represent the retroflex approximant. This alternation will be discussed in greater detail in Chapter 2.

Additionally, capital letters will be used to represent classes of sounds, rather than the specific phones themselves. For example, ⟨C⟩ will be used to represent any stop consonant, as in /sCr/, which is taken to represent all /s{p,t,k}r/ clusters. Similarly, ⟨S⟩ will be used to represent any sibilant /s/, /ʃ/, or intermediate target between the categories. In general, this dissertation assumes that the onset sibilant preceding any consonant other than /r/ is /s/, such that *street* is phonemically /strit/. However, in Chapter 4, I use nonce words with an onset continuum from /s/ to /ʃ/ in attempt to force a contrast between /s/ and /ʃ/. Thus, I will use ⟨S⟩ to refer to the onset sibilant that in those environments may be categorized by

the listener as either /s/ or /ʃ/.

1.3 Outline of chapters

This dissertation is comprised of three experiments that each seek to better understand the phenomenon of /s/-retraction in order to better understand its origin, status, and trajectory. Specifically, this dissertation is structured as follows:

Chapter 2 Prior to presenting the three experiments, I provide an extensive overview to previous research on /s/-retraction, laying the groundwork to motivate the experiments of this dissertation. This includes previous empirical work on the perception and production of the phenomenon, theoretical work on its origins and current grammatical status, and sociolinguistic work to better understand its distribution and potential socio-indexical meaning.

Chapter 3 In Experiment I: Cue Integration, I train participants to associate a given image with a specific word. I then present the participant with an auditory stimulus manipulated to contain a specified degree of /s/-retraction. Using eye tracking in the Visual World Paradigm (Allopenna et al., 1998), this experiment asks if listeners have detailed phonological knowledge about /s/-retraction and, if so, if that make use of that knowledge in real-time speech processing. Specifically, I ask if listeners can use the cues of /s/-retraction to anticipate the presence of an upcoming /r/, correctly looking to a word like *street* before they even hear the /r/.

Chapter 4 In Experiment II: Categorization, I create a continuum from /s/ to /ʃ/, asking listeners to categorize a nonsense word as beginning with an /s/ or an /ʃ/. With this experiment, I ask if listeners account for their expectations of /s/-retraction and adjust their perceptual boundaries accordingly. I examine the age of the listeners to ask how

categorization of these sibilants is changing over time, as younger speakers may have greater experience with /s/-retraction as a sound change in progress. And finally, I examine different individually-defined measures of masculinity, like their evaluation of the faces presented, their evaluation of the model talkers, and their own relative endorsement of traditional stereotypes of masculinity. With these indicators of masculinity, I ask how the potential socio-indexicality of /s/-retraction may influence categorization strategies.

Chapter 5 In Experiment III: Convergence, I use a covert shadowing task to ask if listeners exhibit convergence to manipulated degrees of /s/-retraction. Specifically, I ask if listeners are more likely to converge toward more retracted, i.e. more /ʃ/-like, sibilants, which would increase the relative degree of coarticulation and move the sound change along, or conversely, whether they are more likely to converge toward less retracted, i.e. more /s/-like, sibilants, which would increase the phonological contrast between /s/ and /ʃ/, but impede the transition and propagation of the sound change.

Chapter 6 Finally, I summarize the findings and implications of each experiment in light of the goals of this dissertation and point to the future avenues of research highlighted by this work.

CHAPTER 2

/S/-RETRACTION: PERCEPTION, PRODUCTION, & ORIGINS

As I lay out in the previous chapter, this dissertation investigates the production and perception of /s/-retraction in American English, specifically examining the different patterns that individuals exhibit in cue integration (Experiment I, Chapter 3), phoneme categorization (Experiment II, Chapter 4), and convergence (Experiment III, Chapter 5).

In this chapter, I provide a general overview to /s/-retraction, highlighting previous research that has investigated this phenomenon. In Section 2.1, I begin with a description of the production of /s/-retraction, including its phonological distribution, articulation, and acoustic realizations. In Section 2.2, I turn to the perception and processing of /s/-retraction. In Section 2.3, I present the varying theories for the origins of /s/-retraction and, in Section 2.4, I present the differing theories for the current grammatical status of this sound change. In Section 2.5, I present a brief sociolinguistic sketch of /s/-retraction, focusing on its distribution in apparent time, across the United States and the Anglophone world, and across gender identities. In Section 2.6, I examine the potential socio-indexicality of /s/-retraction, including an overview to previous sociolinguistic research on /s/-retraction and an examination of online meta-commentary surrounding the phenomenon.

2.1 Production of /s/-retraction

In this section, I describe /s/-retraction in detail, which I have briefly introduced in the previous chapter as the process by which /s/ is realized approaching /ʃ/, especially in the context of /r/, such that *street* /strit/ may sound like *shreet* /ʃtrit/. In this section, I first address the phonological distribution of /s/-retraction, that is the phonological environments in which this phenomenon is frequently observed. I then continue to discuss the acoustic and

articulatory realization of sibilants in general and in scenarios which may be characterized as /s/-retraction in particular.

2.1.1 Phonological distribution of /s/-retraction

I have described /s/-retraction as the process by which /s/ is realized approaching /ʃ/, especially in the context of /r/, but as of yet I have not fully explored what it means to be "in the context of /r/". In this section, I discuss in detail the environments environments in which /s/-retraction is most commonly observed in English.

In most dialects of English, /sr/ clusters are phonotactically illicit, such that *shrink* /ʃrɪŋk/ is a well-formed word, but *srink* /srɪŋk/ is not. Conversely, /ʃ/ is phonotactically illicit preceding consonants other than /r/, such that *slim* /slɪm/ is a well-formed word but *shlim* /ʃlɪm/ is not.[1] This preconsonantal neutralization between /s/ and /ʃ/ collapses the phonological contrast between the sibilants made elsewhere. However, through borrowings and phonological changes, /sr/ clusters can emerge and are often subsequently repaired by native speakers of English. For example, in the word *grocery* /groʊsəri/, the post-tonic /ə/ can variably be deleted. This results in an illicit /sr/ cluster, which speakers repair to /ʃr/, yielding *groshry* /groʊʃri/. Similarly, loanwords with /sr/ clusters are often borrowed with an /ʃ/ rather than an /s/, as in *Sri Lanka* /ʃrilɑŋkə/.

Despite being a phenomenon driven by /r/, the most canonical example of /s/-retraction concerns non-adjacent /s/-/r/ sequences. Specifically, /s/ is well-known to retract in /str/ clusters, such that the word *street* /strit/ may be pronounced approaching *shtreet* /ʃtrit/. This phenomenon was first reported by Shapiro (1995) and has generated a significant amount of attention in research on the phonetics, phonology, and sociolinguistics of Englishes around the world. In these environments, the /r/ is generally agreed to exert some influence that yields retraction, although the specific phonological accounts of how that occurs is dis-

1. There are notable exceptions, primarily of Yiddish and German origin, like *spiel* /ʃpil/ and *Schnapps* /ʃnɑps/.

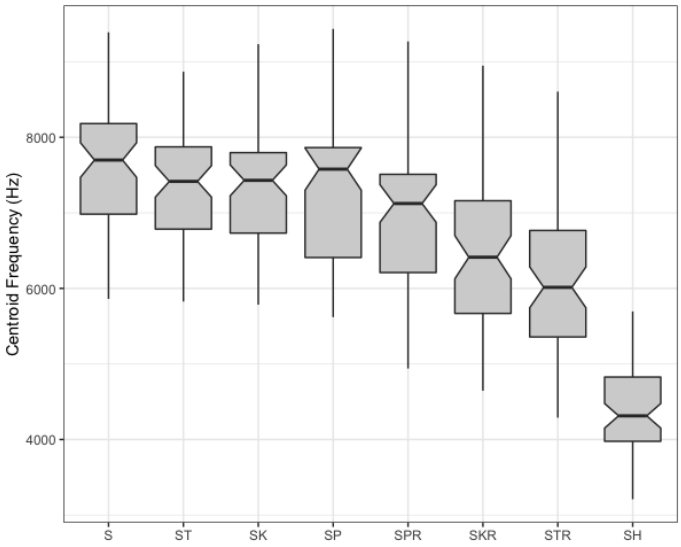


Figure 2.1: Centroid frequency for different sibilant onsets from the pre-test of Experiment III (Chapter 5). A higher centroid frequency (y-axis) indicates a more stereotypically /s/-like sibilant.

puted, which is discussed more in Section 2.3. In fact, this association is so pervasive that for many researchers, /s/-retraction is a phenomenon limited to /str/ environments alone. Why then, if the non-adjacent /r/ is the trigger of retraction, is /s/-retraction not observed in /spr/ or /skr/ clusters? Such that *scream* /skrim/ is rarely pronounced as *shcream* /ʃkrim/ and *spring* /sprɪŋ/ is rarely pronounced as *shpring* /ʃprɪŋ/. Little work has examined the potential for retraction in these clusters or questioned the nature or origin of this asymmetrical distribution.

As it currently stands, it is not the case that /s/-retraction is a categorical process by which /s/ in /str/ clusters is produced identically to /ʃ/, while preconsonantal /s/ elsewhere is identical to prevocalic /s/. In fact, for most individuals, some coarticulatory retraction can be observed in all preconsonantal environments, from minimal retraction in all /sC/ clusters, increased retraction in /spr/ and /skr/ clusters, and the most retraction in /str/ clusters. Figure 2.1 illustrates the varying acoustic realization of sibilants depending on the phonological environment, providing a teaser of the production data from Experiment

III in Chapter 5. I discuss the acoustic metrics used to contrast sibilants and characterize /s/-retraction in the following section, but we can generally characterize /s/ as having a relatively higher centroid frequency and /ʃ/ as having a relatively lower centroid frequency. Thus, as Figure 2.1 demonstrates, /str/ clusters are the most /ʃ/-like, but not categorically so; the articulation of /str/ is still intermediate between /s/ and /ʃ/. Furthermore, /spr/ and /skr/ also exhibit significant retraction, with values intermediate between /sC/ and /str/ clusters. With this distribution in mind, in which there is a clear asymmetry between the places of articulation, but the distribution is by no means categorical, this dissertation examines /s/-retraction as a phenomenon that can be observed in all /sCr/ clusters in order to probe the origins, distribution, and future trajectories of the sound change.

Additionally, in varieties of English that did not undergo post-coronal yod-dropping, in contrast to Standard American English, such that *do* /du/ and *dew* /dju/ are not homophonous, /s/-retraction is also seen in /stj/ clusters. In these dialects, /j/ like /r/ can trigger retraction across the intervening /t/, such that *student* /stjudənt/ may be pronounced approaching *shtudent* /ʃtjudənt/ (Lawrence, 2000; Warren, 2006). As this dissertation is concerned with American English, /stj/ clusters are not examined.

Before delving deeper into /s/-retraction in American English, it is worth noting that similar historic patterns are common cross-linguistically. This is particularly true for other Germanic languages, albeit often in preconsonantal environments more generally and not necessarily in the context of /r/ or /j/. For example, /s/ is produced as /ʃ/ in word-initial clusters in Standard German, as in *Straße* 'street' /ʃtrasə/ and *Spiegel* 'mirror' /ʃpigəl/, and additionally in coda clusters in Swabian German, as in *West* 'west' /vɛʃt/ in contrast to Standard German *West* 'west' /vɛst/ (Bukmaier et al., 2014). Similarly, /s/ is retracted to /ʃ/ preconsonantally in many dialects of Portuguese, as in *pista* 'track' /piʃta/, as well as in similar environments in dialects of Catalan, Italian, Persian, Slovenian, and Spanish, among many others (Kümmel, 2007).

2.1.2 Acoustics

As mentioned in the previous chapter, sibilants are a class of fricatives, which means that they are characterized by aperiodic, turbulent airflow. Sibilants contrast with other fricatives acoustically by having a significant higher intensity (loudness) and frequency (pitch). In fact, sibilants are distinguished from almost all other speech sounds by utilizing, i.e. making contrasts in, a frequency band of their own. In contrast to the perception of vowels and other consonants which generally require cues below 2500 Hz, like formants and formant transitions, sibilants are contrasted by multiple spectral cues above 3000 Hz. The first four spectral moments are four such cues, and result from treating the aperiodic noise as a random probability distribution and analyzing the spectra via Fast Fourier Transforms. The resulting spectral moments are: centroid frequency (M1), or center of gravity, which is an indicator of mean spectral energy, i.e. the frequency band at which the energy is most concentrated; standard deviation (M2), which is an indicator of the range of variance in spectral energy; skewness (M3), which is an indicator of the (a)symmetric distribution of the spectral energy; and kurtosis (M4), which is an indicator of the peakiness of the spectra. Additionally, sibilants can be characterized by their peak frequency, which is the highest peak, i.e. the single frequency with the most energy. No single cue has been found to categorize sibilants between speakers with 100% accuracy (Jongman et al., 2000), but both centroid frequency and peak frequency have been demonstrated to be reliable cues in distinguishing prevocalic /s/ and /ʃ/ generally (Haley et al., 2010; Jongman et al., 2000; Shadle & Mair, 1996, i.a) and /s/ and /str/ in particular (Baker et al., 2011; Rutter, 2011; Smith et al., 2019, i.a). In general, /s/ is characterized by a higher centroid frequency and peak frequency relative to /ʃ/, due to the shorter oral cavity anterior to the constriction necessary for the production of /s/ relative to /ʃ/. Additionally, both spectral measurements have been shown to be highly variable depending on the speaker (Hughes & Halle, 1956), their gender (Nittrouer, 1995; Stuart-Smith, 2007), and socio-economic class (Stuart-Smith, 2007). As /s/-retraction is a

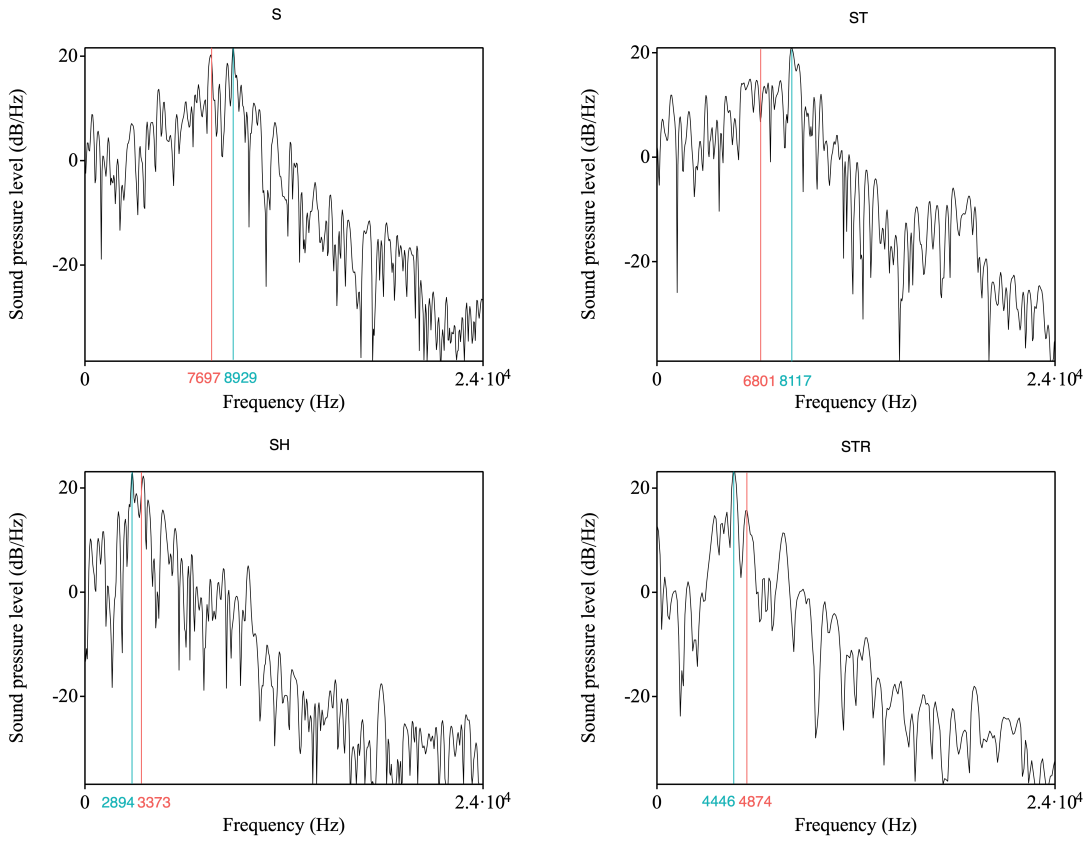


Figure 2.2: Spectral slices for different sibilant onsets (/s/, /st/, /str/, and /ʃ/ clockwise from the top left) from a sample talker (D11), whose /str/ retraction ratio was closest to the community mean. The spectral slice shows the pressure (y-axis) at each frequency (x-axis) for a brief period centered on the sibilant midpoint. Centroid frequency for each onset is indicated in red, peak frequency is indicated in teal.

sound change in which /s/ approaches /ʃ/, a more retracted /s/ is expected to have a lower spectral energy than both a less retracted /s/ and /s/ in non-/str/ environments.

Figure 2.2 provides the spectral slices from the midpoint of four sibilant onsets (/s/, /st/, /str/, and /ʃ/) for a sample talker (D11) from Experiment III (Chapter 5). In each panel, the centroid frequency and peak frequency are indicated, demonstrating that both measurements capture the differences between prevocalic /s/ and /ʃ/, as well as the relative intermediate position of /st/ and /str/ in particular. In order to select the most reliable measurement (centroid frequency or peak frequency) to use in this dissertation, I performed a quadratic discriminant analysis (QDA). Using a subset of the data collected, QDA clas-

sifiers were fit to the Gaussian distributions of prevocalic /s/ and /ʃ/, first using centroid frequency as the cue of contrast. The remainder of the data was then run through through the classifier to determine which distribution (/s/ or /ʃ/) a given measurement most likely belongs to. A second set of classifiers were fit using peak frequency. The QDA classifiers using centroid frequency outperformed those using peak frequency, exhibiting greater accuracy in discriminating prevocalic /s/ and /ʃ/. Following these findings, centroid frequency was selected as the measurement characterizing sibilants for this dissertation.

As I noted before, centroid frequency is highly variable depending on the speaker (Hughes & Halle, 1956) and, as /s/-retraction is the process by which /s/ approaches /ʃ/, an examination of /s/-retraction is less concerned with the raw centroid frequency value than where that value falls compared to the speaker's prevocalic /s/ and /ʃ/. The retraction ratio (Mielke et al., 2010; Baker et al., 2011) is a measurement that accounts for just that, calculating the relative position of a given sibilant between the speaker's mean prevocalic /s/ and /ʃ/. Thus, throughout this dissertation, I will use the retraction ratio, with the formula provided in 2.1, to characterize and measure /s/-retraction. A value of 0 suggests that the given sibilant's centroid frequency value is identical to prevocalic /s/ while a value of 1 suggests that the given sibilant's centroid frequency value is identical to prevocalic /ʃ/.

$$\text{Retraction Ratio} = \frac{\text{CF of segment} - \text{speaker mean CF of /s/}}{\text{speaker mean CF of /ʃ/} - \text{speaker mean CF of /s/}} \tag{2.1}$$

Turning back to the distribution of /s/-retraction presented in Figure 2.1, which uses centroid frequency to characterize the sibilants, Figure 2.3 plots the same data using retraction ratio. With the speaker-normalized values, we again see that /str/ clusters show the most retraction out of all preconsonantal sibilants, with a mean retraction ratio of 0.5, halfway between s/ and /ʃ/. Again, /spr/ and /skr/ exhibit significant retraction, with values intermediate between /sC/ and /str/ clusters.

In contrast to the methods employed in this dissertation, previous research in determin-

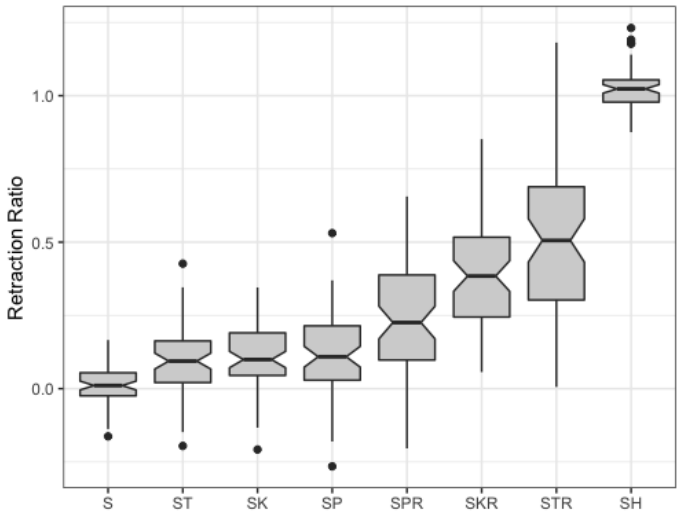


Figure 2.3: Retraction ratio values for different sibilant onsets from the pre-test of Experiment III (Chapter 5). A higher retraction ratio (y-axis) indicates a more retracted, i.e. more /ʃ/-like, sibilant.

ing whether the onset in an /str/ cluster is retracted has typically relied on the researchers' judgments rather than acoustic measurements. However, due to the lack of phonological contrast between /s/ and /ʃ/, it is not demonstrably clear at which point an acoustically retracted /s/ becomes an auditorily retracted /s/. Regardless, rarely does /s/ in /str/ clusters become acoustically identical to a prevocalic /ʃ/; even in instances in which a retracted /str/ is categorized by the researcher as /ʃ/, it is still acoustically intermediate between prevocalic /s/ and /ʃ/.

2.1.3 Articulation

To achieve the high frequency, turbulent noise characteristic of /s/ and /ʃ/, sibilants are articulated by channeling airflow through a narrow constriction and onto the upper teeth. All sibilants are coronal, meaning that they are articulated with the anterior portion of the tongue; the alveolar /s/ and palato-alveolar /ʃ/ differ with exactly which part of the tongue makes the constriction and precisely where that constriction is made. A prototypical /s/ is

apical and alveolar, meaning that the tongue tip forms a narrow constriction at the alveolar ridge, and the tongue is grooved, focusing airflow to that constriction. A prototypical /ʃ/ is laminal and palato-alveolar, meaning that the tongue blade (just behind the tip) makes a narrow constriction just behind the alveolar ridge, and the tongue front (just behind the blade) bunches up along the palate, focusing the airflow toward the constriction. The higher frequency of /s/ relative to /ʃ/ is thus due to the smaller resonating cavity anterior to the constriction, the more precisely directed airflow toward the upper teeth, and the more grooved contour of the tongue, directing and speeding up the airflow before it even reaches the constriction (Stevens, 1988). Furthermore, the palato-alveolar /ʃ/ is typically produced with lip protrusion, by which the lips form an extended, rounded shape that serves to lengthen the anterior constriction, lower the frequency, and thus enhance the contrast between /s/ and /ʃ/.

For /s/-retraction, regardless of the phonological model proposed, the trigger for the sound change, either directly or indirectly, is /r/. Unlike sibilants, in which small changes in lingual position or contour can have outsized acoustic impacts, English /r/ can be produced with wildly different tongue postures with no acoustic effect (Delattre & Freeman, 1968). This many-to-one articulatory-to-acoustic mapping has lead to a system of covert variation, by which inter- and intraspeaker variation is observed in /r/ articulation without any perceptual consequences (Mielke et al., 2016). The two most common articulations of /r/ are as bunched /ɹ/ or retroflex /ɻ/ (Mielke et al., 2010). In a bunched articulation, the tongue tip is drawn down and the tongue body bunches up near the palate. In a retroflex articulation, the tongue tip is raised and curled back. In an examination of American English speakers, Mielke et al. (2016) demonstrate that half of the speakers produce a bunched articulation exclusively, while half use a retroflex in at least some phonological environments. Furthermore, /r/, like /ʃ/, is typically produced with some degree of lip protrusion, which is also found to differ widely both within and between speakers, varying from no rounding, to outward

protrusion, to vertical compression (Delattre & Freeman, 1968; Mielke et al., 2016).

Many of the proposals described in greater detail later in the next section rely on one articulation of /r/ or another as the basis of /s/-retraction. In this vein, Mielke et al. (2010) hypothesize that the *covert*, i.e. not perceptible, variation in /r/ articulation might lead to *overt* variation in /s/-retraction. Mielke et al. find that not one of the participants produced a retroflex /ɻ/ in /str/ clusters, which was the only environment in which tongue posture uniformity was observed. Building off this observation, Baker et al. (2011) examine the same speakers separately either as 'retractors' or 'non-retractors' based off the researchers' categorization of the speakers. Baker et al. find that for individuals who are not perceived as retractors, the degree of retraction was predicted by the similarity in tongue shape between the bunched /ɹ/ and /s/. Thus, for individuals who were perceived as producing an /s/ in /str/ clusters, the more similar the tongue shape between /s/ and /r/, the greater the retraction ratio of, i.e. the more /ʃ/-like, the onset sibilant. No such correlation was observed for more categorical retractors. Thus, for individuals who were perceived as producing an /ʃ/ in /str/ clusters, tongue shape similarity between /s/ and /r/ had no bearing on the observed retraction ratio. Smith et al. (2019) expand this by training a classifier to categorize ultrasound frames of the tongue and video stills of the lips from /str/ clusters as either the rhotic /ɹ/ or the palato-alveolar consonants /ʃ, t͡ʃ, d͡ʒ/. They find that, across the board, the onset sibilant in /str/ clusters looked more palato-alveolar than rhotic both lingually and labially. What this classifier crucially does not ask is whether the onset sibilant in /str/ clusters is more likely to be categorized as /s/ or /ʃ/, as the acoustic evidence illustrates that even for the most retracted individuals, the onset sibilant in /str/ clusters is still intermediate between prevocalic /s/ and /ʃ/. No research to my knowledge has examined the articulation of /spr/ or /skr/ clusters in detail.

2.2 Perception of /s/-retraction

While the production of /s/-retraction has been the focus of a growing body of linguistic research in recent years, the perception of the phenomenon has received much less attention. Only two studies to my knowledge have examined the perception of /s/-retraction. Both studies focus on how sibilants are categorized in prevocalic environments and what that can tell us about the perception (and production) of /s/-retraction.

Before diving into the studies examining /s/-retraction, let us first consider the perception of sibilants in general. Sibilants, like many other speech sounds, are generally perceived categorically (Feldman et al., 2009; Fujisaki & Kawashima, 1969; Mann & Repp, 1980; Repp, 1981, i.a.). Categorical perception (Liberman et al., 1957) is the perception of distinct, and often abrupt, categories when presented with gradient changes to speech sounds along a continuum. That is, the potentially discriminable differences between the speech sounds are perceived as belonging to the same class until suddenly a category boundary is approached and they are perceived as belonging to separate, independent classes. Thus, for sibilant sounds like /s/ and /ʃ/, a gradient continuum can be created along an acoustic measurement like centroid frequency, yet listeners will perceive only two distinct categories: /s/ and /ʃ/. The category boundaries vary between listeners and are adjusted depending on the speaker and contextual information, including the phonological environment. In an example presented in greater detail in Chapter 4, Mann & Repp (1980) find that listeners compensate for the expectation of coarticulation in speech perception. That is, listeners may perceive an identical sound differently in different phonological environments, as they factor out the coarticulatory noise in order to arrive at the intended target. For sibilants, Mann & Repp show that listeners compensate for coarticulatory lip rounding from the following vowel, which by lengthening the cavity anterior to the constriction may make /s/ sound more like /ʃ/. Thus, the same sibilant step on a continuum from /s/ to /ʃ/ may be perceived as /ʃ/ when preceding an unrounded vowel, but as /s/ when preceding a rounded vowel.

Turning to the influence of /s/-retraction on sibilant categorization, Kraljic et al. (2008) ask whether listeners may adjust their category boundaries between /s/ and /ʃ/ depending on the source to which they attribute they variation they are presented with. In New York/Long Island English, where /s/-retraction is particularly advanced, Kraljic et al. exposed participants to one of two systems: a context-conditioned system in which the speaker exhibits a dialectal pattern (i.e. /s/-retraction in /str/ clusters like in *industry* or *artistry*) and a context-indepedent system in which the speaker exhibits an idiolectal pattern (i.e. /s/-retraction in arbitrary /sV/ sequences like in *dinosaur* or *medicine*). Participants were first exposed to this model talker and then completed a category identification task for a [asi]–[aʃi] or [astri]–[aʃtri] continuum. Kraljic et al. found that participants exposed to the idiolectal variants show significantly greater shifts from the pre- and post-exposure tasks in /s/-/ʃ/ categorization than participants exposed to the dialectal variants. The results suggest that identical segments can have different impacts on the perception of phonological contrasts depending on their phonological environment. Whether this finding has to do with factors such as dialectal association and familiarity or coarticulation and phonetic naturalness is not clear from the experiment design. Additionally, Kraljic et al. examined participants' production, including an explicit imitation task, which is discussed in greater detail in Chapter 5.

In Australia, where /s/-retraction is reportedly less advanced than the United States, Stevens & Harrington (2016) examine the perception of /s/-retraction in /str/ clusters, which the authors suggest exhibit the preconditions for the sound change. Stevens & Harrington first elicited production data from Australian speakers, including instances of prevocalic /s/ and /ʃ/, as well as /s/ preceding /t/ and /tr/. In the subsequent perception study, the initial sibilant from *seam*, *sheep*, *steam*, and *stream* were crossed spliced onto –/it/. Listeners were then presented with a forced choice between *seat* and *sheet*. Each instance of phonological /ʃ/ was categorized correctly as /ʃ/, i.e. *sheet*, by all listeners; however, some

listeners categorized instances of phonological /s/ from either *seam* or *steam* as /ʃ/ and all listeners categorized at least one instance of phonological /s/ from *stream* as /ʃ/. These results suggest that even in a community described as having only the pre-conditions for /s/-retraction, the native production of /s/ in /str/ clusters by some speakers was retracted enough to be perceived as belonging to a different category.

Both of these studies suggest that listeners may have context-specific perceptual strategies for /s/-retraction without testing the categorization of these environments specifically. In part, this is a result of English phonotactics, making testing categorization inherently challenging when only one phoneme is a licit response. In Experiment II: Categorization (Chapter 4), I introduce nonce words in order to force a contrast between /s/ and /ʃ/ pre-consonantally. With this experiment, I build off these studies to ask whether listeners are more categorical or more continuous in the perception of /s/ in these environments where a true contrast does not exist. Furthermore, both of the studies presented above demonstrate that listeners have some knowledge of and familiarity with /s/-retraction. However, it is unclear whether this information is useful to listeners. That is, how does that knowledge influence speech processing and does it aid the listener in making decisions? In Experiment I: Cue Integration (Chapter 3), I ask if listeners can use that knowledge in real-time speech processing to aid lexical activation and prediction. Such a finding would not only demonstrate that /s/-retraction is a process that is beneficial to the listener as well as the speaker, but also it would support a path to sound change by which the cue weight shifts toward the onset /s/, potentially catalyzing the push from coarticulation to categorical change.

2.3 Theories regarding the origins of /s/-retraction

In the previous sections of this chapter, I have presented the empirical results of studies on the perception and production of sibilants in general and /s/-retraction in particular. In the present section, I highlight the research that has sought to better understand the origins of

the sound change itself. The variety of methods employed by researchers, along with the different speaker communities and subject pools, have contributed to multiple phonological and phonetic accounts of /s/-retraction. Many of these accounts are not mutually exclusive, but rather highlight a different aspect of the sound change or a different perspective of the researcher or researchers. Furthermore, many of these accounts rely on the impressions and intuitions of the researcher, rather than using acoustic or articulatory data, and thus may present proposals that contradict empirical evidence that has since been made available. When relevant, I discuss how new evidence supports or challenges these accounts.

2.3.1 Coarticulation

As mentioned in the previous chapter, coarticulation is the phenomenon by which a sound is influenced by neighboring sounds. Many proposals of /s/-retraction suggest that the sound change emerged due to the significant coarticulatory influences of /r/. In this section, I examine proposals that coarticulation of the tongue and lips could have led to the sound change.

Lingual coarticulation

As its name suggests, /s/-retraction is often primarily thought of as concerning the retraction of the tongue. In general, /r/ is produced with a tongue posture that is significantly more retracted relative to a typical /s/. Thus, from a coarticulatory standpoint, /s/-retraction is the process by which the tongue posture becomes more similar to the upcoming /r/, in order to minimize the articulatory distances between /s/ and /r/ and thereby reduce articulatory effort. By minimizing the articulatory distance between /r/ and /s/ and retracting the tongue, then acoustic result is onset sibilant that sounds more /ʃ/-like.

Baker et al. (2011) propose that /s/-retraction is a prototypical example of coarticulation as the precursor to sound change. Using ultrasound imaging of the tongue, Baker et al.

examine tongue postures for speakers' production of both /s/ and /r/, finding evidence for clear coarticulatory effects in tongue posture. Strikingly, the effect is stronger for the individuals who do not yet produce an onset sibilant perceived by the researchers as /ʃ/. Baker et al. interpret these findings to suggest that their participants provide a window into the two different states of /s/-retraction: the precursors to change and categorical sound change. It is in the precursor stage that the coarticulatory effects are strongest, with a more /ʃ/-like sibilant produced with a smaller articulatory distance between /s/ and /r/.

The question remains then, how do we go from a state of coarticulation to sound change? And, given that the non-adjacent /s/-/r/ sequence existed long before /s/-retraction as sound change emerged, why did it emerge at all? Research in historical linguistics has long proposed that the systemic variation that is a consequence of coarticulation may lead to categorical sound change. For example, Paul (1888) proposed that small articulatory deviations can lead to change if they "incline to one side or the other" rather than vary in a random distribution. The predominant view today is that at some point a listener misinterprets an aberrant instance of coarticulation as the intended target and adjusts their production accordingly (Beddor, 2009; Ohala, 1993, i.a.). This crucially requires that individuals have different grammars regarding the state of the phenomenon and that over time a preponderance of individuals' grammars are modified from coarticulation to phonologization (Beddor, 2009).

However, Gylfadottir (2015) suggests that an account of /s/-retraction as coarticulation is potentially problematic as it cannot explain the observed asymmetries between /s{p, t, k}r/ clusters. An account of retraction as coarticulation from /r/ to /s/ would not capture the differences in rates of retraction between /str/ and /s{p, k}r/ clusters, as the overlapping gestures of /r/ and /s/ would not be hindered anymore by /p/ or /k/ than they would be by /t/. In fact, /p/, with no intervening lingual gesture, should offer the strongest potential environment for /s/-retraction. However, these asymmetries do not in and of themselves discredit an account of retraction as coarticulation. Rather, they suggest that the intervening

consonants must play some critical role in this process. Furthermore, some coarticulation is in fact observed in /spr/ and /skr/ clusters, just not to the same degree as /str/ clusters. Thus in a coarticulatory path to /s/-retraction, varying degrees of coarticulation may be predicted in all /sCr/ clusters, but the intervening consonant may play some role in the subsequent phonologization of the sound change. Whether the nature of that role is articulatory, acoustic, perceptual, or phonological has yet to be determined.

Labial coarticulation

As briefly mentioned earlier, /s/ and /r/ differ not just in tongue posture, but also in the position of the lips. Specifically, /r/, like /ʃ/ is produced with some degree of lip protrusion, while /s/ typically is not. Thus, in the same way that coarticulation between sounds contributes to tongue backing, it also contributes to lip protrusion. And like tongue retraction, lip protrusion leads to a more /ʃ/-like sound.

While /s/-retraction undoubtedly involves a retraction of the tongue body, it is possible that coarticulatory lip protrusion played a critical role in the origin of the sound change. Specifically, this would explain why /s/-retraction is least observed in /spr/ clusters, as the labial closure in /p/ may block coarticulatory lip protrusion. It is possible that coarticulatory lip protrusion chronologically preceded lingual coarticulation and provided an acoustic effect of retraction that encouraged later tongue retraction. Alternatively, it is possible that both coarticulatory processes arose simultaneously, or even that lingual coarticulation chronologically preceded labial coarticulation, but that the coarticulatory lip protrusion emphasized the acoustic effect of tongue retraction in such a way to catalyze the sound change and facilitate phonologization in /str/ clusters alone. However, these hypothesis remains largely untested.

2.3.2 Assimilation at a distance

Initial phonological accounts of /s/-retraction, including Shapiro (1995), view this sound change as a case of long distance assimilation, by which /s/ assimilates the place features of /r/ across the intervening /t/. Shapiro states that /s/-retraction is a phonemic change by which /s/ is palatalized, but "the degree of palatalization is not uniform, so that the phonetic realization can stop short of the full-fledged 'phonetic power' of the American [ʃ] found in words like *short*, *shape*, *ash*, etc." (p. 101). That is, Shapiro proposes that /s/ phonemically becomes /ʃ/ in /str/ clusters, or rather /ʃtr/ clusters, but does not become phonetically identical to prevocalic [ʃ]. Under such an account, there is still context-induced variation between prevocalic and preconsonantal sibilants, but the phonological features of the sibilant in /str/ clusters must change to the most similar phonemic category. Specifically, this account suggest that /s/ in /str/ clusters assimilates the [+ compact] feature of /r/, which would now generally be described as [− anterior]. The relatively simple phonological change proposed by Shapiro is thus presented in 2.2.

$$/\text{str}/ \longrightarrow /\text{ʃtr}/ \tag{2.2}$$

Shapiro (1995) contends that this /s/-retraction must be a long distance process, as there is no phonemic change in /t/ that would locally induce assimilation. Specifically, Shapiro states that since /t/ in /str/ clusters is not aspirated, as stops in /sC/ clusters in American English are not typically aspirated, /t/ cannot affricate[2]. This contrasts with the /t/ in initial /tr/ clusters, where /t/ is certainly aspirated and often affricated (see next section for details on affrication). Thus, according to Shapiro, /t/ undergoes no phonemic change to affrication, despite its increased VOT in /str/ clusters compared to /st/ clusters prevocalically. As such, "/t/ remains /t/ before /r/ after /s/ no matter what the phonetic

2. Later work, including Smith et al. (2019) discussed later in this section, suggest that /t/ can and does affricate /str/ clusters for some speakers.

characteristics of [t] are here", and thus /t/ cannot "figure into the assimilation of /s/ to /r/" (p. 103). Without a change in the phonemic nature of /t/, Shapiro contends that the only possible explanation of retraction is the distant assimilation of /s/ to /r/.

If /s/-retraction is treated as a case of long distance assimilation in which the intervening consonant plays no role, this predicts that phonemic changes should be observed in all /sCr/ clusters. Shapiro acknowledges this observation and accounts for the differences between /str/ and /spr/ clusters by proposing "that /str/ lends itself to realization as [ʃtr] while /spr/ remains unaltered [...] due to the phonological value of /t/ and /p/" (p. 104). While not explicitly mentioned, one can assume that the same argument holds for /skr/ clusters as well. Specifically, this accounts suggests that [+ acute] feature of /t/, which would now generally be described as [+ coronal], permits long distance assimilation across the stop, while the [− acute] feature of /p/ blocks it. Another way of describing this alternation would be appealing to underspecification: Coronal stops like /t/ may be unspecified or underspecified for place, and thus the place features of /r/ can spread across them to /s/, but /p/ and /k/, which are fully specified for place, block that feature spreading. Such an account would predict that no retraction, or at least no categorical change, is observed in /s{p, k}r/ clusters.

2.3.3 Local assimilation

In a direct response to Shapiro (1995), Lawrence (2000) contends that characterizing retraction as assimilation at a distance is misguided and is better explained following strictly local theories of assimilation. Lawrence proposes that /s/-retraction is still motivated by /r/, but realized by the local effect that /r/ has on /t/. In the present section, I examine three local processes of assimilation, including spectral lowering affects of /t/, affrication of /t/, and retroflexion of /t/.

Coronal spectral lowering

A completely local account of /s/-retraction would appeal to the general lowering effect that /t/ has on the centroid frequency of /s/. While no work to my knowledge has proposed this account, let us consider it briefly as a straw man argument before turning to local assimilation processes that are ultimately driven by the non-adjacent /r/. Empirical work, including the foreshadowed data from Experiment III: Convergence presented Figures 2.1 and 2.3, has found that there is a general lowering effect of /s/ preconsonantally. Various researchers (Baker et al., 2011; Mielke et al., 2010; Stevens & Harrington, 2016) have found that /s/ is observed to have a lower centroid frequency, and thus is more /ʃ/-like, when followed by /{p, t, k}/, with an even greater lowering effect in /s{p, t, k}r/ clusters. While the empirical findings presented earlier find no place difference between the /sC/ clusters, Mielke et al. (2010) and Baker et al. (2011) find that /s/ is significantly more retracted in /st/ clusters than /sp/ and /sk/ clusters. However, such a strictly local account, by which /s/ is retracted preceding /t/, cannot explain why retraction is categorically more advanced in /str/ clusters than /st/ clusters.

Affrication

Another local account of /s/-retraction is affrication, in which the initial trigger of retraction is still /r/. In these accounts, first proposed by Lawrence (2000), /s/ locally assimilates the place features of the /t/ in /str/ clusters. In order for this process to occur, /t/ must first affricate to /t͡ʃ/ preceding /r/.

Stops are known to be released more slowly preceding approximants, especially /r/, thus increasing the relative duration of aspiration stop-approximant clusters (Sievers, 1901; Klatt, 1975). As the stop constriction is released, airflow begins through the narrow constriction produced for the approximant, leading to increased turbulence, which can then be misinterpreted as intentional affrication (Ohala & Solé, 2010; Hall et al., 2006). Thus, /tr/

targets may be perceptually reanalyzed as /t͡ʃr/ clusters. The same pattern is observed in voiced coronal stops, with /d/ affricating to /d͡ʒr/. While aspiration duration also increases prerhotically for bilabial and velar stops, the VOT difference is greater in coronal-rhotic clusters, and the affrication reanalysis is unique to those clusters (Ohala & Solé, 2010; Stevens, 1971). Additionally, examination of children's 'invented spellings' provides a unique window into this reanalysis, with /t/ in /tr/ clusters frequently written as ⟨CH⟩, such that children write *chruck* in lieu of *truck*, and /d/ in /dr/ clusters as ⟨J⟩, such that children write *jrink* in lieu of *drink* (Read, 1975; Treiman, 1985). Smith et al. (2019) provides empirical evidence for affrication in /tr/ and /dr/ as a sound change in progress, both in perceptual and acoustic metrics, with affrication increasing in apparent time and speakers born after 1980 approaching ceiling. Furthermore, Smith et al. (2019) observe that the articulatory similarity between /t/ preceding /r/ and prevocalic /t͡ʃ/ is particularly striking for the youngest speakers, suggesting that the sounds have been phonologically merged.

Although Olive et al. (1993) and Shapiro (1995) posit that /t/ should not affricate in /str/ clusters since /t/ is not aspirated when following a sibilant in English, Olive et al. (1993) also note that the VOT in /s{p, t, k}r/ clusters is still significantly longer than in /s{p, t, k}/ clusters due to following approximant. In fact, Olive et al. (1993) note that there is often no consistent abrupt cessation in frication that would typically characterize a stop in these clusters. Lawrence (2000) notes that he perceives /t/ to be affricated in retracted /str/ clusters in New Zealand English, leading him to posit that /s/-retraction is not caused by assimilation at a distance, but rather is the result of local assimilation of adjacent features, as /s/ assimilates the palato-alveolar place features of the affricate. Thus, /t/ affricates in the presence of /r/ due to perceptual and articulatory motivations, which subsequently motivates the retraction of /s/ to /ʃ/ as a local assimilatory process, as modeled in (2.3).

$$/\text{str}/ \longrightarrow /\text{st}\widehat{\int}\text{r}/ \longrightarrow /\widehat{\int\text{t}}\int\text{r}/ \tag{2.3}$$

This is parallel to the retraction of /s/ that is also frequently observed before [tʃ] in Standard American English as in *moisture* or *question*, as schematized in (2.4).

$$/\text{st}\widehat{\int}/ \longrightarrow /\int\widehat{\text{t}\int}/ \tag{2.4}$$

This account captures the proposed asymmetry in retraction, predicting retraction in /str/ clusters and not in /s{p, k}r/ clusters. However, this account does not account for the frequent moderate shifts in centroid frequencies in these clusters, nor does it predict the rare categorical shifts that have been observed for some individuals (Phillips, under review).

Smith et al. (2019) examine /tr/ affrication, /dr/ affrication, and /s/-retraction to better understand the relationship between the processes. While they find a significant correlation between speakers who affricate /tr/ and /dr/ onsets, there is no such correlation for individuals who affricate /tr/ onsets and retract /str/ onsets. However, an implicational hierarchy appears to emerge: That is, an individual who affricates /tr/ is not significantly more likely retract /str/, but there are no individuals who retract /str/ but do not affricate /tr/. While not direct evidence for an affrication path to retraction, this does not rule out such an account. Furthermore, as mentioned in Section 2.1, Smith et al. (2019) also collected articulatory data and trained a classifier to categorize ultrasound frames of the tongue and video stills of the lips from /str/ clusters as either the rhotic /ɹ/ or the palato-alveolar consonants /ʃ, $\widehat{\text{t}\int}$, $\widehat{\text{d}\text{ʒ}}$/. They find that not only was /s/ more likely to be classified as palato-alveolar than rhotic, but so too was the intervening stop. Smith et al. propose that this finding illustrates that a phonological change to /$\widehat{\text{t}\int}$/ has taken place, rather than just potential coarticulation from /r/. However, there was no classifier option for /t/, which would permit the model to suggest that no change, whether coarticulatory or phonological, was observed.

Retroflexion

Another possible local process involves retroflexion rather than affrication. Under such an approach, /s/ approaches a retroflex articulation /ʂ/ rather than a palato-alveolar /ʃ/. In such an account, briefly suggested but impressionistically discarded by Shapiro (1995), the intervening /t/ assimilates the retroflex place features of the rhotic, and in turn /s/ assimilates those retroflex features from /t/, now /ʈ/, which is schematized in (2.5).

$$/\text{st}\underset{\frown}{\text{ɹ}}/ \longrightarrow /\text{s}\underset{\frown}{\text{ʈɹ}}/ \longrightarrow /\underset{\frown}{\text{ʂʈɹ}}/ \tag{2.5}$$

Models like in (2.5) may explain the asymmetry between the retraction that occurs in /str/ clusters but not in /spr/ and /skr/ clusters by appealing to the underspecificity of coronals, thus allowing /t/ to more easily assimilate the place features of /ɹ̠/. Alternatively, such accounts may assert that the assimilatory distance between alveolar and retroflex is much less than that between velar/bilabial and retroflex, and thus a more natural sound change. Furthermore, this would predict that retraction is only observed, or at least only originated, with speakers who select retroflex articulations of /r/. However, while Shapiro (1995) notes the possibility of a retroflex assimilatory model, he suggests that perceptually he heard palato-alveolar [ʃ] and not retroflex [ʂ] in /str/ clusters. This tentative hypothesis was confirmed by recent studies (Mielke et al., 2010; Baker et al., 2011; Archangeli et al., 2011).

Mielke et al. (2010) used ultrasound imaging to determine the tongue shape of the sibilants and rhotics in the different phonological environments for speakers of American English. For speakers who vary in their selection of retroflex or bunched /r/, retroflex articulations were generally disfavored preceding coronals and front vowels, while they were favored next to antagonistic segments including labials, word boundaries, and back vowels. This distribution can be explained by phonetic naturalness: With all else being equal, the articulation of /r/ with the shortest articulatory distance from the adjacent phones is selected. The results

for /s/-retraction are particularly striking: Despite the variation observed in the selection of bunched or retroflex articulations of /r/ in different positions, only bunched /r/ is observed in /str/ clusters for all participants. However, while Mielke et al. (2010) find higher rates of retroflexion in /sp/ and /sk/ clusters than in /st/ clusters, no results are reported for rate of retroflexion in /spr/ and /skr/ clusters. If we assume that even a single instance of retroflexion can be observed in /spr/ and /skr/ clusters, there is not the same uniformity of bunched articulations as was observed for /str/ clusters.

The difference in rhotic selection in these clusters may help explain the asymmetries in rates of retraction. If retraction were to occur in all clusters as a form of assimilation to the rhotic, this hypothesis predicts differences in sibilant articulation across /s{p, t, k}r/ clusters, i.e. a palato-alveolar sibilant preceding a bunched /r/ – [ʃtɹ], and a retroflex sibilant preceding a retroflex /r/ – [s̪pɻ]. The asymmetries in observed retraction rates may suggest that retroflex tongue shapes are assimilated at lower rates than bunched tongue shapes.

The observed asymmetries in retraction may additionally be explained by the possible inconsistency in rhotic selection in /spr/ and /skr/ clusters. While no retroflexion was observed in /str/ clusters, it is possible that it can be observed in /spr/ and /skr/ clusters. However, it is unlikely that it would be observed in all instances of such clusters, especially in words like *spree* or *scream*, as bunched articulations are preferred preceding front vowels regardless of the preceding consonant. The non-uniformity of rhotic articulations in these clusters may explain the lack of consistent retraction and its unlearnability. Mielke et al. (2010) discuss rhotic selection as the 'Peter Pan' of sound changes: It is unlearnable and thus will never reach completion because of its inconsistent and idiosyncratic distribution. However, there is at least one environment in which there is consistent and non-idiosyncratic selection: the bunched selection in /str/ clusters. While consistency of rhotic selection itself is not a motivation for the sound change, if bunched articulations facilitate retraction, the consistency of rhotic selection may lead to consistency in an individual's and a community's

/s/-retraction production, which in turn may allow for that sound change to be learned and phonologized. In contrast, the possible inconsistency of rhotic selection in /spr/ and /skr/ clusters can explain why retraction in these clusters is less observed.

2.4 Theories regarding the current grammatical status of /s/-retraction

In the previous section, I entertained different proposals for the origins of /s/-retraction as a sound change, from coarticulatory to assimilatory models. In the present section, I discuss the current status of the sound change, asking whether categorical sound change has occurred or whether the current situation is better characterized as coarticulatory variation.

One potential indicator of the current grammatical status is the distribution of /str/ onsets between canonical /s/ and /ʃ/. In contrast to proposals that /s/-retraction is a categorical process with few intermediate forms between /s/ and /ʃ/ (Rutter, 2011), most acoustic analyses of /s/-retraction have observed intermediate forms for /str/ clusters (Baker et al., 2011; Durian, 2007; Gylfadottir, 2015; Labov, 2001; Mielke et al., 2010; Phillips, under review; Smith et al., 2019). While little previous work examines /s/-retraction in /spr/ and /skr/ clusters, intermediate forms between /s/ and /ʃ/ have also been observed in those environments (Mielke et al., 2010; Phillips, under review), including the teaser data from this dissertation included in Figure 2.3. It is worth noting that the observation of intermediate forms does not preclude that the sound change has been phonologized for some or all speakers. This is because regardless of the production of a sibilant in these environments, the phonological contrast between /s/ and /ʃ/ will remain neutralized in that environment. Context-induced variation may be observed for the preconsonantal sibilants regardless of their phonological identity as either /s/ or /ʃ/. Furthermore, it is possible that the sound change may be phonologized to a new category and distribution entirely. In their agent-based modeling of /s/-retraction, Stevens et al. (2019) find greater prediction

for a split versus a merge phonologization of /s/-retraction. In this account, /str/ clusters split from prevocalic /s/ and move toward, but never merge with, /ʃ/, instead forming a third sibilant category intermediate between the two. Nonetheless, a preponderance of /str/ tokens identical to or at least very similar to prevocalic /ʃ/ would provide evidence for the phonologization of the sound change.

Baker et al. (2011) propose that, in its current state, /s/-retraction can be observed acting as both a coarticulatory and phonological process. Baker et al. propose that, for individuals categorized as retractors, i.e. individuals perceived to produce /ʃtr/ clusters, /s/-retraction can be seen as a categorical, phonological process. While for individuals categorized as non-retractors, i.e. individuals still perceived to produce /str/ clusters, /s/-retraction can be seen as a gradient, coarticulatory process. This is supported by the differences between the two groups in the correlation of acoustic retractedness and tongue posture similarity between /str/ and /r/. Non-retractors show a strong correlation, suggesting evidence for coarticulation, while no such correlation is observed for retractors, suggesting evidence for phonologization. Additionally, working with the same subject pool, Mielke et al. (2010) observed an apparent gap in retraction ratio values, with almost no individuals exhibiting a retraction ratio between 0.6 and 0.8. Nearly all individuals below 0.6 were categorized as non-retractors, and all above 0.8 were categorized as retractors. Neither Mielke et al. nor Baker et al. explicitly address individual variation in /spr/ and /skr/ clusters. In a recent continuation of this research, Smith et al. (2019) find that, in the few years since the publication of Mielke et al. and Baker et al., phonologization is occurring rapidly in the Raleigh area, with most individuals now exhibiting a retraction ratio closer to 1, i.e. an /str/ onset nearly identical to /ʃ/.[3]

Findings like Baker et al. (2011) and Mielke et al. (2010) suggest that /s/-retraction can be thought of as both a coarticulatory and phonological process simultaneously. This is in

3. This may be further aided by a change in their calculation of retraction ratio, using /st/ instead of prevocalic /s/ in normalization.

part because of differences between individuals who may have different grammars and because the differing approaches capture and characterize distinct aspects of the phenomenon. Many of the other descriptions of /s/-retraction presented thus far have typically selected either a coarticulatory or phonological account and dismissed a possible role of other factors in the phenomenon. On one side, Shapiro (1995) suggests that the phonetic realization of /t/ cannot influence the assimilatory process of /s/. Similarly, some work in coarticulation has dismissed an influence with phonology, suggesting that it is a universal and biomechanical process, and thus the phonology of an individual or a language should not influence coarticulatory processes. However, previous work has suggested language-specific effects of coarticulation (Manuel, 1990) and demonstrated that coarticulation is planned (Whalen, 1990; Solé, 2007) and varies by communicative setting (Lindblom, 1990; Lindblom et al., 2007; Scarborough, 2013). The language-specific effects of coarticulation identified by Manuel suggest a role of the phonological system in governing coarticulation. Thus, a treatment of /s/-retraction as coarticulation need not dismiss phonological influences and can integrate them into a coarticulatory model.

For this dissertation, I continue to treat /s/-retraction as a primarily coarticulatory, rather than assimilatory, process. This is motivated in large part by the fact that the majority of the speakers recruited for this dissertation appear to exhibit retraction ratios that suggest coarticulation rather than categorical phonologization. When relevant, I discuss how individual differences between the participants of this experiment may be explained by differentiating between phonologized retractors and coarticulatory non-retractors.

2.5 Sociolinguistic distribution of /s/-retraction

Beginning with Shapiro's (1995) inaugural study of /s/-retraction, a growing body of researchers have noted the unique sociolinguistic profile that /s/-retraction has among sound changes. In the present section, I focus on three aspects of the sociolinguistic distribution

/s/-retraction. First, I demonstrate that it is a sound change in progress with trajectories in apparent time (Section 2.5.1). Second, I examine its wide-spread geographic distribution without a single apparent origin (Section 2.5.2). Thirdly, I turn to its gender distribution and the differing findings therein (Section 2.5.3). In the following section, I examine the potential socio-indexical role of /s/-retraction.

2.5.1 Apparent time studies

Any sound change in progress that is truly in progress should exhibit clear differences over time. Sociolinguists examine time trends through two lenses: real and apparent time. Real time studies are conducted longitudinally and recruit large, intergenerational subject pools to observe the change as it progresses. Apparent time studies examine the speech of different generations at one point in time to ask how the language may have changed. With the latter being logistically much more feasible, apparent time research relies on the hypothesis that older generations represent an encapsulated form of the language at a previous stage and that any variation between generations is not due to stable age-grading or physiological senescence. Three notable examinations of /s/-retraction have used multi-generational sociolinguistic interviews to investigate apparent time changes in /s/-retraction: Durian (2007) for Columbus, Ohio, Gylfadottir (2015) for Philadelphia, Pennsylvania, and Wilbanks (2017) and Smith et al. (2019) for Raleigh, North Carolina.

Durian (2007) conducted 32 sociolinguistic interviews with speakers from the Columbus area aged 19 to 69. Durian extracted the first ten instances of /str/ for each speaker and categorized the sibilant as either [s], [ʃ], or 'intermediate', impressionistically or with the aid of the spectrogram when necessary. Using three age bins (19-29, 30-49, 50-69), Durian finds that that proportion of [ʃ] or 'intermediate' ratings decreases as the subject age increased, suggesting an apparent time finding of increased retraction for Columbus, Ohio.

For her investigation, Gylfadottir (2015) uses the Philadelphia Neighborhood Corpus

and the Influence of Higher Education on Local Phonology, both large existing corpora of sociolinguistic interviews with speakers in the Philadelphia area. Between the two corpora, Gylfadottir examines 225 speakers born between 1888 and 2004. Looking at the birth year of her subjects, Gylfadottir finds that /s/-retraction has advanced significantly in the Philadelphia area over the generations examined.

Similarly, Wilbanks (2017) and Smith et al. (2019) examine 162 speakers from Raleigh, North Carolina, born between 1918 and 1996, from a corpus of sociolinguistic interviews. Over the span of time represented by the speakers, Raleigh had undergone significant urbanization and increased immigration from the northeastern United States. Smith et al. (2019) find that, for the female speakers, /s/-retraction has advanced rapidly and relatively abruptly in word-medial position, as in *restaurant* or *constraint*, with a more gradual increase in word-initial positions. For the male speakers, /s/-retraction remains relatively stable in apparent time in both word-initial and medial positions.

2.5.2 Geographic distribution

Shapiro (1995) describes /s/-retraction as a "general American innovation", as it doesn't appear to be tied to any particular region but can be described as more eastern than western, more southern than northern, and more rural than urban. Since then, a growing body of work has demonstrated that /s/-retraction is a wide-spread phenomenon, although its prevalence and acoustic realizations vary from region to region. In research conducted in our laboratory, we have observed a retracted /s/ in /str/ clusters for various speakers across the continental United States (Phillips, under review). Regional specific analysis for the United States have observed speakers exhibiting retraction from the South (Rutter, 2011 for Louisiana; Phillips, 2001 for Georgia; Wilbanks, 2017 and Smith et al., 2019 for North Carolina; Hinrichs et al., 2016 for Texas), the Midwest (Durian, 2007 for Columbus), and the Northeast (Labov, 1984 and Gylfadottir, 2015 for Philadelphia; Kraljic et al., 2008 for New York). Research from the

Western United States found that /s/-retraction is much less advanced in Arizona, California and Washington than the Eastern United States (Mielke et al., 2010 and Baker et al., 2011).

The phenomenon of /s/-retraction is not limited to the continental U.S., as research has noted observations of /s/-retraction robustly in New Zealand (Lawrence, 2000; Bauer & Warren, 2004; Warren, 2006), as well as many other parts of the Anglophone world including Newfoundland (Clarke, 2004) and England (Glain, 2013, 2014; Cruttenden, 2014; with Altendorf, 2003 specifically for Estuary English and Bass, 2009 for Cochester English). Additionally, Stevens & Harrington (2016) found evidence for precursors to /s/-retraction in Australia, although, like the the Western United States, it is not as advanced as in other regions.

2.5.3 Gender distribution

Among his principles of linguistic change, Labov (2001, p. 280) generalizes that "women have been found to be in advance of men in most of the linguistic changes in progress studied by quantitative means" for changes from 'below', where 'below' characterizes sound changes below the level of consciousness and of lower prestige. As Gylfadottir (2015) and Hinrichs et al. (2016) suggest that /s/-retraction is below the level of consciousness and does not carry overt prestige (see next section), at least in Philadelphia and Austin respectively, one would expect that women would exhibit more /s/-retraction than men. However, in her cross-generational examination of Philadelphia English speakers, Gylfadottir (2015) finds that /s/-retraction is not more advanced in female speakers than male speakers. In contrast, research from Raleigh, North Carolina finds no differences between men and women in perceived retraction, where the onset sibilant is auditorily coded by the researchers as 'retracted' or 'non-retracted', but finds that younger women lead men acoustically, both in centroid frequency and retraction ratio (Smith et al., 2019; Wilbanks, 2017). On the other hand, in examinations from the U.K., men were found to use /s/-retraction more than women (Bass,

2009; Glain, 2014).

The observed gender differences between Philadelphia, Raleigh, and the U.K. are not necessarily contradictory, as they may be indicative of differences in the socio-indexical meaning of /s/-retraction between the regions. Specifically, /s/ in general (i.e. prevocalic environments) has been demonstrated to be robustly linked to gender and gender typicality in both production and perception (Campbell-Kibler, forthcoming; Podesva & Van Hofwegen, 2014; Strand, 1999; Zimman, 2017, i.a.). As the socio-indexical role of sibilant variation in preconsonantal environments has not been explored extensively (see next section), it's possible that it is associated with gender performance in one region but not the other, e.g. it *may* index masculinity in the U.K., but index urban and northern alignment in Raleigh. It's also possible that from an auditory perspective, /s/-retraction is less notable for female speakers whose sibilants are typically higher frequency, and thus more stereotypically /s/-like, than men. In the following section, I examine the potential socio-indexicality of /s/-retraction, including the potential role of /s/-retraction in indexing masculinity and the link between /s/ and gender/gender typicality.

2.6 Potential socio-indexical meaning of /s/-retraction

As mentioned in the previous section, few examinations of /s/-retraction have specifically looked at its potential socio-indexical role(s). In this section, I first provide an overview to research on the socio-indexicality of sibilants in general, including in prevocalic environments. I then present the previous work on socio-indexicality of /s/-retraction, including studies from which socio-indexical meaning is inferred from variation within a community and studies in which individuals are made aware of the sound change and explicitly asked for a response. Finally, I examine online meta-commentary surrounding /s/-retraction in order to better understand how the way in which non-linguists are talking about /s/-retraction can inform our understanding of its potential social meaning.

2.6.1 Prevocalic or context-independent socio-indexicality of sibilants

Many acoustic analyses of English have robustly found that male speakers consistently produce /s/ with a lower centroid/peak frequency than female speakers (Fuchs & Toda, 2010; Sachs et al., 1973; Podesva & Van Hofwegen, 2014). Similarly, male speakers with more traditional gender identities, histories, and sexualities produce /s/ with a lower centroid frequency than less norm-conforming men (Linville, 1998; Podesva & Van Hofwegen, 2014; Zimman, 2017). Trans-men were observed to produce similar centroid frequency values to cis straight men and significantly lower values than cis gay men (Zimman, 2013).

In speech perception, sibilants have been proposed to be a salient, available social index, often linked to sexuality or perceived sexuality, with a fronted /s/ perceived as more gay for a perceived male talker (Crist, 1997; Linville, 1998; Levon, 2007; van Borsel et al., 2009; Munson, 2010; Mack & Munson, 2012, a.o.). These studies, however, crucially examine sibilant fronting, which is often perceived as the stereotypical 'gay lisp'. Less work, however, has examined the socio-indexical meaning of sibilant backing.

In one notable exception, Campbell-Kibler (2011a) examines the perception of backed, fronted, and mid-range /s/ in her investigation of the acoustic contribution to perceived sexuality in male speakers. Campbell-Kibler examines all three variant productions together, as her previous research (Campbell-Kibler, 2011b) illustrates that two sociolinguistic variants can have distinct social associations. In her analysis of *-in/-ing* variation, she finds that listeners perceive a speaker presented with *-in* to be more casual, but a speaker with *-ing* is not perceived as less casual. Similarly, a speaker presented with *-ing* is perceived as more intelligent and more educated, but a speaker presented with *-in* is not perceived as less intelligent or less educated. Turning toward variants of /s/ production (examined in tandem with *-in/-ing* variation), Campbell-Kibler (2011a) finds that listeners perceive a fronted /s/ (the 'gay lisp') as more gay than a mid-range and backed /s/ across all speakers, but a backed /s/ is not necessarily perceived as less gay than a mid-range /s/. Rather,

a backed /s/ can contributes a more 'country' identity than a fronted or medial /s/ for a southern speaker. While competence and masculinity are generally linked, with voices perceived as more masculine also perceived as more competent, a backed /s/ examined in conjunction with the *-in* variant can be perceived as both less competent and more masculine, indexing a 'masculine, unintelligent, straight man' style. Conversely, a fronted /s/ examined in conjunction with *-ing* can be perceived as more competent and less masculine, indexing a 'sassy gay friend' persona.

Following the findings of backed sibilants indexing rural identity, Podesva & Van Hofwegen (2014) examine sibilant production in a rural community in California. They find a stronger polarization between the sibilants of men and women in the area, highlighing an acoustic realization of the social conservatism of the area. Likewise, Podesva & Van Hofwegen (2014) find that a retracted /s/ not only indexes an alignment toward the country, but also heteronormative masculinity and non-heteronormative femininity. In other work on sibilant backing, Stuart-Smith (2007) observed class differences in sibilant production in Glasgow, with working-class women producing /s/ with a lower centroid frequency than middle-class women.

Taken together, these different studies on the perception and production of prevocalic /s/ demonstrate that it is a complex indicator than can have complex interactions with other variants. However, on the whole, a more retracted /s/ is repeatedly tied to more stereotypically masculine identities.

2.6.2 Preconsonantal socio-indexicality of sibilants

Specific examinations of the social meaning attributed to /s/-retraction in /sCr/ clusters is much more limited. In some instances, social meaning in inferred from the sociolinguistic distribution of the sound change. For example, from the sociolinguistic interviews examined, Gylfadottir (2015) suggests that there is little evidence to demonstrate that /s/-retraction

has reached the level of consciousness in Philadelphia, with a few anecdotal accounts of speakers expressing their disapproval of the variable, including one elementary school speech therapist. From sociolinguistic interviews in the Columbus, Ohio area, Durian (2007) found that /s/-retraction, including tokens categorized as intermediate and [ʃ], is more prevalent in speakers raised in Columbus than the surrounding suburbs, and that its use has increased in the suburbs over time. Durian notes that these findings coincide with the urbanization of the greater Columbus area and changes in migration patterns (from city-to-suburbs to suburbs-to-city). Ultimately, Durian suggests that /s/-retraction in Columbus may serve to index a general urban affiliation, even among speakers from the suburban areas.

Using a rapid anonymous survey, Hinrichs et al. (2016) first elicited /str/ clusters from individuals on the street in Austin. After marking a production as /s/ or /ʃ/, Hinrichs et al. explicitly explained the phenomenon to their participants and asked for their intuitions about whom they would associate with a /ʃ/-like onset. Hinrichs et al. found that many speakers were unaware of the phenomenon, suggesting it is below the level of consciousness. When prompted to give social meaning to the phenomenon, speakers provided diverse responses, often suggesting that it may be more 'Southern/hick', non-native, or the result of a speech impediment. However, as the previous section has demonstrated, the reality of /s/-retraction is both prevalent and native.

Recent work has also sought to answer this question, appealing to the implicit connections listeners make between /s/-retraction and social attributes rather than naming the phenomenon and explicitly asking for listeners' impressions. Phillips (2018) asked eight undergraduate students at the University of Chicago to read a series of sibilant initial words, including /s/, /spr/, /str/, /skr/ and /ʃ/ onsets. The eight model talkers showed varying degrees of retraction in /str/ clusters, with one male and one female exhibiting forms I perceived to be /ʃ/ rather than /s/, i.e. the retractors.

Each of the sibilant initial words was segmented and embedded in an open-ended panel

study on Amazon Mechanical Turk and completed by 15 native speakers of American English with internet connections in the United States. Panel participants heard the unmanipulated sibilant-initial stimuli for each model talker and were asked to supply any information about the speaker's identity, characteristics, or traits in an open-ended question. For the male retractor, panel participants were split in describing his /str/ tokens as more masculine, athletic, and straight than his other /s/ tokens, or whether those same tokens were more gay, intellectual, and pretentious. The female retractor was described as younger in her /str/ onsets compared to her other /s/ tokens. Following the panel study, seven attributes were selected for the subsequent social evaluation task, including attractiveness, masculinity, friendliness, formality, shyness, and sexuality. Additionally, geographic regions and environment types were included.

The sibilant target words were then manipulated to contain varying degrees of retraction; prevocalic /s/ and /ʃ/ were first digitally mixed and subsequently cross-spliced onto the target words. 342 native speakers of American English participating with internet connections in the United States were recruited on Mechanical Turk. Participants listened to the manipulated stimuli and rated the talker on the seven attributes listed above. Phillips (2018) found that not only do individual participants vary significantly in how they rate the model talkers, but also individual talkers vary consistently in how they are rated. For speakers who are perceived as more typical for a given trait, such as straightness or masculinity, there is no observed difference between the clusters and no clear association between retraction and these traits. However, for speakers who are less typical on a given trait, the associations of retraction expected for prevocalic sibilants hold, albeit weakly, in /spr/ and /skr/ clusters. As illustrated in Figure 2.4, a male talker who was generally perceived as more gay than the other male talkers, was perceived as significantly straighter in /skr/ and /spr/ clusters in more retracted, i.e. less /s/-like, conditions. However, there was no observed difference in his /str/ evaluation as a result of retraction condition, which is precisely where /s/-retraction is

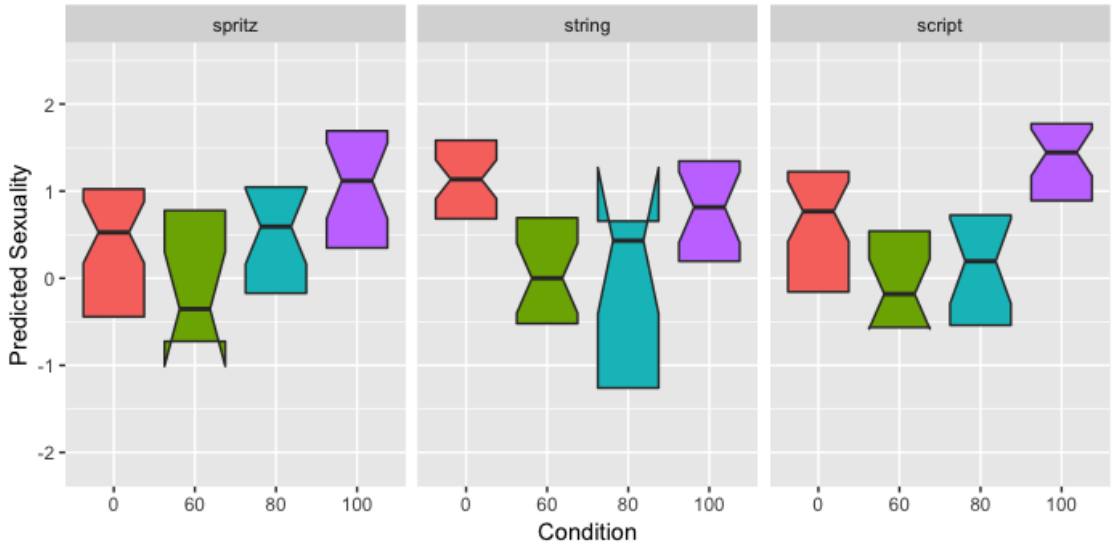


Figure 2.4: Results from Phillips (2018) illustrating the predicted sexuality in the social evaluation of a model talker rated more homosexual than his peers. Predicted sexuality (z-scored) is plotted by word (panel) and percent /s/ onset (0-100, x-axis). A positive predicted sexuality value indicates a more homosexual response, while a negative value indicates a more heterosexual response.

most expected. These findings challenge extending the expected socio-indexicality of prevocalic sibilants to preconsonantal environments in general and to /str/ clusters specifically.

2.6.3 Online meta-commentary surrounding /s/-retraction

The findings from both a rapid anonymous survey (Hinrichs et al., 2016) and social evaluations task (Phillips, 2018) are inconclusive and point to a high degree of individual variability and uncertainty, suggesting that more work is needed to better understand the potential socio-indexical work of /s/-retraction. In addition to the meta-commentary elicited from speakers in the rapid anonymous survey on the street, Hinrichs et al. (2016) examined meta-discourse of /s/-retraction in online forums. Like the respondents in Austin, individuals online demonstrate incredibly varied social attributions. Some individuals attribute /s/-retraction to the Northeastern U.S., specifically the New Jersey shore and Long Island, while others view it as more southern or hick; some posters view it as associated with

women, femininity, or gayness, while others associate it with men, masculinity, and straightness. However, as none of these findings have been published and their sources cannot be confirmed, I replicated this search to examine what the meta-discourse around /s/-retraction is online and how that can inform our understanding of its social meaning.

The first challenge of examining meta-commentary surrounding /s/-retraction is finding instances of it to begin with. The methodology adopted in this section was to conduct a Google search for a modified spelling of high frequency words with /str/ clusters, for example *shtreet* instead of *street*. Each link from the first ten pages was consulted for *street*, the highest frequency word with an /str/ cluster according to Subtlex (Brysbaert & New, 2009, see Chapter 3.3.1 for details). For the second through tenth highest frequency /str/ words (*straight, strong, strange, destroy, strike, district, struck, strip* and *instructions*), the first five pages of hits were consulted, but for many of these words fewer than five pages of hits were available. Additionally, a Twitter search for each of the modified spelling as a hashtag (e.g. #shtrong) were consulted. Through this initial search, it became evident that one lower frequency word repeatedly popped up in the discourse on /s/-retraction: *struggle*. Thus, additional Google and Twitter searches were conducted for *struggle*. The complete list of online sources consulted and their relevant details is provided in Appendix A.3.

The results of the search support previous suggestions that /s/-retraction remains below the level of consciousness for most speakers, with many posters sharing that they only recently noticed it, had never noticed it before reading a previous post, or are frustrated that no one else is aware of /s/-retraction to the point that others tell them they are mistaken. While the majority of the posts expressed disapproval of the sound change, there were many posts and forums devoted to better understanding the sound change from a hobby linguist point of view, including the sharing of many of the academic publications cited in this dissertation. Furthermore, while the majority of the comments were limited to /str/ environments, only one post mentioned /spr/ clusters, noting that a commercial for *Sprite*

Table 2.1: Common meta-discourse associations and their frequency count

Communities		Traits		Places		People	
Black	27	Speech disorder[1]	17	Philadelphia	12	Michelle Obama	30
German	13	Drunk	8	New York	9	News anchors[2]	25
Non-native[3]	7	Tough/masculine	7	Northeast	9	Sean Connery	11
Jewish	6	Hood/street	4	Kerry, Ireland	9	George W. Bush	6

[1]including lisps, dentures, orthodontia, and craniofacial anomalies
[2]including radio, television, news, and sports reporting
[3]including general immigrant references, but excluding specific nationalities, e.g. German

sounded like *Shprite*; no comments were found discussing /skr/ clusters. Many disapproving comments extended discussion to [ʃ] in other environments they thought unwarranted by the word's spelling, despite being the standard pronunciation, like *sure* or *location*.

Additionally, the posts reflect the previously reported breadth of diverse associations between /s/-retraction and traits, groups, people, and places, with the most frequent associations included in Table 2.1. The most common associations were to black communities and talkers, especially Michelle Obama. However, even within this category, there was a tremendous degree of variability in what /s/-retraction is indexing in the black community, spanning associations of a hood or street persona ("I always thought it was a 'ghetto' thing, like 'axe' for 'ask' or 'A-ight' for 'alright"') to an educated black persona ("It seems that when Blacks used this pronunciation back then it signaled an educated person"). The associations with Michelle Obama are particularly noteworthy because they outnumber all other associations. In part, the attention Ms. Obama receives is due to the relative size of her platform, but the number of her mentions dwarfs those of George W. Bush. But it is also worth noting that in 2016, Rush Limbaugh publicly imitated and mocked Ms. Obama's pronunciation of *struggle*, after which mentions of Ms. Obama increased significantly in searches for *shtruggle* and with particularly more ire. However, in many forums, mentions of Ms. Obama predate Mr. Limbaugh's on-air criticism.

In addition to Ms. Obama and Mr. Bush, other politicians (e.g. Paul Ryan, Bernie

Figure 2.5: Example images of online meta-commentary on /s/-retraction.

Sanders, and Sarah Huckabees Sanders) and journalists (from NPR radio to ESPN, see figure 2.5) are repeatedly associated with /s/-retraction. Other individuals with high profiles like actors, athletes, and other celebrities, attract less attention – Sean Connery is the obvious exception, yet he is notable for producing an [ʃ]-like /s/ across the board, even prevocalically. The focus on political figures and journalists may suggest that /s/-retraction is less expected in more formal or standarized speech ("I first noticed this in some speakers of Black American English and George W. Bush, but when the announcer on public radio did it, I got upset. People look to public radio for correct standardized American pronunciations."). It's unclear whether individuals associate /s/-retraction with news speech or whether these associations point to the opposite: /s/-retraction may be implicitly indexed as casual or uneducated, causing listeners to notice it more in political and news speech because that is where it is least expected. That is, listeners may be making adjustments for retractions in casual conversations but not when listening to the evening news.

Given the previous work on prevocalic sibilants, it is perhaps surprising to see less discussion of masculinity and /s/-retraction. There were far fewer associations with masculinity, and the properties of straightness and toughness, than with black communities or the northeastern United States. However, when there were, the comments often spoke more to narrow character traits and social performances than macrodemographic categories and locations.

These associations were found almost exclusively in searches for *shtraight* and *shtrong*. On Twitter, the use of #shtrong seems to index a gym rat persona and is frequently used to describe a tough workout ("Work out wit no spikes on a dirt track.. Thats how you get #Shtrong"). For *shtraight*, this unsurprisingly is used in discussion of sexuality and its interface with masculine toughness. One poster expressively describes how they interpret the socio-indexicality of /s/-retraction in popular culture:

> When movie characters [...] use it, I think they are saying: Only pussies and mamma's boys would say straight when you can say shtraight. Straight is for fastidious librarians; shtraight is for guys who know how to load a gun and stuff a 20 down a shtripper's g-string. It's an anti-lisp. It says: Not only am I not gay, but I'm almost unbelievably shtraight.'

This association is particularly noteworthy given the tenuous association found by Phillips (2018) between masculinity, straightness, and /s/-retraction in a social evaluation task.

Figure 2.6 is an approximation of an indexical field for /s/-retraction, following Eckert (2008). Figure 2.6 is based off the online meta-commentary and thus includes the various, diverse associations that individuals report online. Following Eckert, different associations are represented differently orthographically, not to 'distinguish between two distinct categories of meanings, but to emphasize the fluidity of such categories and the relation between the two in practice' (2008, p. 469). In a break from Eckert, I have also chosen to represent the relative frequency of the associations orthographically, given the diversity of responses. This list is exhaustive, as any association with more than two separate reports is included, but it is curated, such that associations are positioned near other associations viewed by the researcher to be related. This approach, while subjective and heavily influenced by the researcher's own biases, allows for associations to cluster around different traits or social types. This clustering may point to potential social meaning that is otherwise obfuscated when each association is considered independently, especially in cases like the present where the sheer diversity of responses can be overwhelming.

From Figure 2.6, five distinct but often overlapping clusters of associations emerge. Two of these cluster around immutable, non-performative traits: the speech impediment cluster

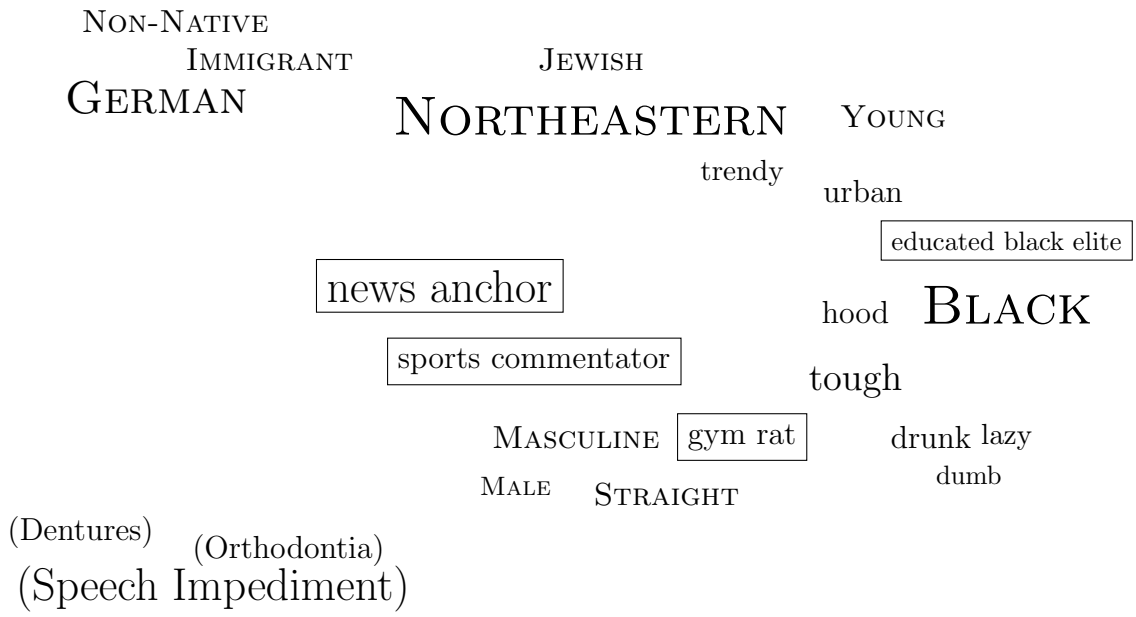


Figure 2.6: Indexical field for /s/-retraction from online meta-commentary, following (Eckert, 2008). Social types are boxed, permanent qualities are in small caps, and traits are in roman weight. No stances, i.e. moods, emotions, etc., were reported. The relative size of the label corresponds to the frequency of the association.

and the non-native cluster. Both of these potentially point to the /s/-retraction status as below the level of consciousness for most individuals (Recall that even many of the posters online state that they have are not familiar with it or that the original poster must be mistaken). That is, even individuals who are explicitly aware of /s/-retraction view it as something aberrant, foreign, and unintentional. The remaining three clusters point to the urban northeastern United States, black communities, and masculinity. Of note, the associations clustering around black communities do not point to a single social type, but rather to the diversity within the black community. For instance, one individual associates it with the educated black elite while another with a ghetto persona. In contrast, the associations clustering around masculinity do not highlight the diversity of social types among male identifying individuals, but rather focus on the intersection of masculinity, sports, and toughness. Taken together, these associations point to the gym rat persona, a man who subscribes to masculine norms of toughness and who, at times obsessively, focuses on manifesting mas-

culinity through physical strength.

Although the reported associations with the northeastern U.S. and the black community outnumber the reported associations with masculinity and toughness, these are precisely the associations that point to a socially meaningful, performative, and narrowly defined persona. Furthermore, these findings mirror and build off the known socio-indexicality of backed sibilants generally. For this reason, insofar as this dissertation seeks to better understand the social meaning conveyed through /s/-retraction, I will focus on how /s/-retraction can index masculinity and, in particular, masculine stereotypes of toughness.

CHAPTER 3

EXPERIMENT I: CUE INTEGRATION

This dissertation examines the perception and production of /s/-retraction, an ongoing sound change in American English by which /s/ approaches [ʃ] in /str/ clusters (such that *street* may sound like *shtreet*) due to the coarticulatory influences of /r/, but less so in /spr/ and /skr/ clusters (*scream* rarely sounds like *shcream*). The aim of the first experiment of this dissertation is to ask whether listeners attend to /s/-retraction in their perception of /sCr/ clusters. That is, do listeners have detailed phonological knowledge about /s/-retraction as a phenomenon and adjust their perceptual strategies accordingly? Specifically, this experiment asks whether individuals can make use of the spectral cues of retraction in real time in speech processing. With /s/-retraction framed as a process of anticipatory coarticulation from the trigger /r/ to the target /s/ across an intervening consonant /C/ (see Chapter 2.3 for more on this account), I ask if listeners are able to use that anticipatory coarticulation to better predict if there is an upcoming /r/. Specifically, when given a forced choice between an /sCr/ word like *string* and an /sC/ word like *sting*, I ask if listeners are able to identify the /sCr/ word faster and more accurately if the onset sibilant contains a significant degree of retraction. Using eye-tracking throughout this process, I examine precisely when listeners use the cues of retraction, asking whether listeners may anticipate the presence (or absence) of an /r/ and execute looks toward the correct word even before the /r/ is heard. Furthermore, I examine the three intervening stops /p, t, k/ to ask whether listeners exhibit different perceptual patterns for one place of articulation compared to the others, they may have more experience with /s/-retraction in /str/ clusters and thus have a potential greater expectation for meaningful context-dependent variation in those clusters compared to other places of articulation.

In this chapter, I first introduce the relevant research on cue integration (Section 3.1), highlighting research that has focused on integration of anticipatory coarticulation and re-

search that examines cue integration strategies for sibilants. I then present an overview to the experiment (Section 3.2) and move on to introduce the materials, procedure and hypotheses of this experiment (Section 3.3). Next, I present the results of the experiment (Section 3.4) and conclude with a discussion of the findings and their implications for the role of anticipatory coarticulation for the listener (Section 3.5).

3.1 An introduction to cue integration

Coarticulation has often been considered to be a process that aids the speaker, as it reduces the articulatory distance between two gestures, thereby potentially decreasing articulatory effort (Lindblom, 1990). Conversely, it has been suggested that coarticulation is a process that inherently hinders the listener by rendering the speech signal more ambiguous (e.g. Stevens & Keyser, 2010). Specifically, some accounts propose that in listener-directed speech, speakers will minimize coarticulation in order maintain or enhance contrasts, thereby increasing articulatory effort, but consequentially increasing listener intelligibility (Lindblom, 1990). However, different lines of research have found weak or no evidence for reduced coarticulation in elicited clear speech (Matthies et al., 2001; Scarborough, 2004).

Conversely, some researchers propose that coarticulation, while not enhancing contrasts for individual phones, provides the listener with helpful contextual information from the overlapping phones, potentially easing the perception of the sounds in their relevant contexts. This perceptual benefit of coarticulation has been included in various phonological theories, including gestural theories (Fowler, 1996) and TRACE models of speech perception (Elman & McClelland, 1986). Additionally, the perceptual benefit of coarticulation has been demonstrated in the laboratory. In lexical processing, individuals are able to correctly identify the target word more quickly and accurately when more coarticulatory information is present (Martin & Bunnell, 1981; Whalen, 1991; Connine & Darnieder, 2009, i.a.). In other tasks, listeners are better able to identify deleted segments when coarticulatory information

is present than when it is missing (Ostreicher & Sharf, 1976).

The development of eye-tracking and mouse-tracking, and the addition of these methodologies to traditional lexical processing tasks, has allowed researchers to get not only response accuracy, but also fine-grained temporal information about when that decision is made. These innovations have led to recent studies which demonstrate that listeners use coarticulatory information in real time, correctly identifying the target from coarticulatory information preceding the contrastive sound before that sound has even been heard. Beddor et al. (2013) examined the perception of nasal coarticulation, presenting the listener with two pictures that varied only on the presence or absence of the nasal consonant, e.g. *scent* /sɛnt/ and *set* /sɛt/. Participants were first familiarized with the set of images, learning to associate an image of a nose smelling some flowers, for example, with *scent* and a grouping of chess pieces with *set*. The results found that listeners were consistently able to direct their gaze toward the word with a nasal consonant, e.g. *scent*, after only hearing the nasalized vowel, before the nasal consonant itself was perceived. Contrariwise, the absence of nasality on the vowel did not lead to faster or more accurate looks toward words without a nasal consonant, e.g. *set*. These findings demonstrate that the coarticulatory 'noise' can be advantageous to some listener who use that coarticulatory information as soon as it is available.

The process by which listeners immediately use available information in lexical identification has been referred to as immediate integration or a 'cascade' perception strategy. Immediate integration been observed in a variety of different environments in which multiple cues may be present sequentially. For example, voiced and voiceless stops are primarily distinguished by voice onset time, but vowel length is a secondary cue. Research in cue integration has demonstrated that lexical processing is immediately sensitive to voice onset time, but can be updated as vowel length information becomes available (McMurray et al., 2008; Toscano & McMurray, 2012; Reinisch & Sjerps, 2013; Kingston et al., 2016).

In contrast to a cascade or immediate integration strategy, a buffer strategy of cue integration suggests that listeners wait until all relevant cues are available before beginning lexical identification. Recent work by Galle et al. (2019) has suggested that listeners use a buffer strategy for the perception of sibilants. That is, despite the potential for listeners to use spectral cues to accurately distinguish prevocalic /s/ and /ʃ/, listeners wait until the vowel onset in lexical identification, holding the spectral information in a 'buffer' until the formant transition information is available. Galle et al. explore a variety of different explanations as to why listeners use a different auditory processing strategy for sibilants than other speech sounds:

- *Cue reliability*: While the spectral energy is more reliable than the formant transition, listeners have experience with potential mismatches between the spectrum and formant transition, such that an /s/ spectrum might have an /ʃ/ transition. This may require them to wait until the formant transition, although it's a secondary cue, to begin lexical identification.

- *Context*: Sibilants are highly context-dependent such that this may require listeners to wait until the vowel in order to compensate for coarticulation or account for talker differences. However, immediate integration has been observed for vowels which are also highly context-dependent.

- *Auditory grouping*: Sibilants are different from most other sounds as they are highly aperiodic and utilize higher frequency bands than other sounds. Perhaps formant transitions are necessary to group the sibilant with the following vowel and categorize it a speech sound at all.

- *Vowel-as-organizer*: Speech may be processed using the vowel as the organizer. As such, lexical activation cannot begin before the vowel or vocoid is perceived. While immediate integration has been observed for stops this may be due to their relatively

shorter duration than sibilant. Alternatively, it may be the case that voice onset time is not truly integrated until vowel onset.

An additional hypothesis explored by Galle et al. (2019) is that the artificially generated sibilants created for their experiment may lack critical information present in naturally-produced sibilants. For that reason, listeners may select a buffer strategy not utilized in natural speech processing. To test this, Galle et al. conducted a gating task, finding that spectral information in the frication alone was sufficient to accurately categorize the artificially produced fricatives. Galle et al. confirm this by examining lexical identification with naturally-produced sibilants, again finding that listener employ a buffer strategy in lexical identification of sibilants. These findings suggest that the artificial nature of the stimuli cannot be responsible for the buffer strategy that appears to be unique to the processing of sibilants, and the observation that listeners wait for formant transitions to integrate spectral information results from one of the aforementioned reasons.

The present experiment examines fricative cue integration preconsonantally where formant transitions are not an available cue. The question here is not what cues distinguish /s/ from /ʃ/, i.e. a phonological contrast between the two sibilants, but what cues distinguish /SC/ from /SCr/, i.e. the presence or absence of an upcoming sound. Thus, I ask whether, like with vowel nasalization, individuals can make use of coarticulatory spectral cues to anticipate the presence or absence of /r/ in lexical processing.

3.2 Study overview

Experiment I: Cue Integration asks if listeners use the coarticulatory cues of /s/-retraction in lexical identification as soon as those cues are available. Approaching /s/-retraction can be viewed as a coarticulatory process, I ask whether listeners begin to expect an upcoming /r/ from retraction on the onset sibilant alone, using a cascade strategy of speech perception, or whether they wait until the onset of the /r/, using a buffer strategy.

In this experiment, participants are presented with two images that contain different phonological environments that may condition the realization of the sibilant, for example /str/ in *string* vs. /st/ in *sting*. Using a modified Visual World Paradigm (Allopenna et al., 1998), participants receive auditory instructions to select one image or the other. All the while, their eye movements are recorded, providing a window into their decision-making on which image to select in real time. Eye movements are an ideal measurement of speech processing because they provide a high temporal resolution of when different lexical representations, or 'candidates', are considered.

There are three conditions for this experiment: decreased, increased, and hyper-increased retraction. In the decreased retraction condition, the onset sibilant in all /sCr/ clusters is manipulated to be more /s/-like and have a higher centroid frequency, indicative of a less retracted /s/, than the model talker's average production for that cluster. In the increased retraction condition, onset sibilants in /sCr/ are manipulated to appear more retracted than typical. In the hyper-increased retraction[1], onset sibilants were manipulated to be further increased, approaching a canonical, prevocalic /ʃ/. This allows for a comparison not just between /sCr/ and /sC/ clusters, but also between decreased, increased, and hyper-increased retraction conditions within the same cluster.

In Section 3.3, I outline the methods and materials used including the stimuli creation (3.3.1), participants and procedure (3.3.2), and analysis and hypotheses (3.3.3). In Section 3.4, I present the results of this experiment, and, in Section 3.5, I move onto a discussion of the findings and their implications.

1. The hyper-increased retraction condition was added post-hoc to the experiment design following preliminary examination of the results from Experiment III: Convergence run concurrently with Experiment I: Cue Integration. Those findings, explored more in Chapter 5, found that many participantins' baseline retraction ratios were higher, i.e. more /ʃ/-like, than the increased retraction stimuli. One goal of the increased retraction condition was to expose subjects to values sufficiently higher than their own. Since the original two conditions were counterbalanced between-subjects, additional subjects were recruited for the hyper-increased retraction condition.

3.3 Methods & materials

3.3.1 Stimuli

The stimuli for this experiment are designed to manipulate the degree of retraction in sibilant clusters to examine whether the anticipatory cues of /ɹ/ presence can influence lexical processing. The stimuli thus not only consist of the relevant /sCr/ clusters, but also /sC/ clusters and simplex prevocalic /s/ and /ʃ/ to serve as competitors. Stop-initial stimul are also included in order to serve as fillers.

All stimuli, both sibilant- and stop-initial, are single English lexical items containing containing the desired onset followed by /ɪ/. Whenever possible, lexical items with coda consonants were selected. Except when it was impossible to complete the relevant paradigm, all stimuli are monosyllabic.

All auditory stimuli, both sibilant- and stop-initial, consisted of naturally-produced lexical words with manipulated onsets. Both classes of stimuli were recorded in the carrier phrase: 'Now select X'. A college-aged male from Illinois was recruited to read each phrase five times. All stimuli materials were recorded at 48,000 Hz with a Shure SM10A head-mounted microphone in a sound-attenuated booth.

Sibilant-initial stimuli The sibilant-initial stimuli contained prevocalic /s/, /ʃ/, as well as sibilants in /sC/ and /sCr/ clusters, across bilabial, alveolar and velar places of articulation. This yields a four-item target paradigm for each place of articulation, e.g. for bilabial: *sit*, *spit*, *spritz*, and *shit*. The relative infrequency of many of the /sCr/ clusters leads to imperfect paradigms where the coda occasionally varied minimally between the near-minimal pairs. However, as the target segments are in the word onset, variations in the coda should not influence looks to either image during the time windows examined. The complete list of the sibilant-initial stimuli is included in Table 3.1, grouped by the onset. Phonemic representations are included and crucially assume that all sibilants preceding consonants are

	word	IPA	$SUBTL_{WF}$	lg10WF
/s/	sip	/sɪp/	5.10	2.42
	sing	/sɪŋ/	97.59	3.69
	sit	/sɪt/	311.35	4.20
/sp/	spit	/spɪt/	19.35	2.99
/spr/	spritz	/sprɪts/	0.49	1.42
/st/	sting	/stɪŋ/	7.02	2.55
/str/	string	/strɪŋ/	12.67	2.81
/sk/	skip	/skɪp/	21.1	3.03
/skr/	script	/skrɪpt/	19.61	3.00
/sh/	ship	/ʃɪp/	98.88	3.70
	shingle	/ʃɪŋgəl/	0.75	1.59
	shit	/ʃɪt/	474.65	4.38

Table 3.1: Sibilant-initial wordlist for Experiment I: Cue Integration

/s/.

Lexical frequency for each lexical word in the experiment is included in Table 3.1. These values were taken from $SUBTLEX_{US}$ (Brysbaert & New, 2009), a corpus of American English that derives its frequency counts from a corpus of over 51 million words collected from the subtitles of 8,388 American films and televisions episodes from 1900-2007, with the majority post-1990. The television shows and films that comprise the corpus come from a variety of genres with the intent of representing lexical frequency of American English in general, across dialects, styles, and registers. $SUBTL_{WF}$ is the word frequency per million words and lg10WF is the logged measurement of the count of a given word in the corpus + 1.

Prior to manipulating the sibilant onsets, I examined the sibilants naturally-produced by the model talker, measuring the centroid frequency of /s/ in the different onset environments to inform the manipulations. Time-averaged centroid frequency values were extracted for

/S/	CF	RR	/sC/	CF	RR	/sCr/	CF	RR
/s/	6439	0.00	/sp/	6185	0.11	/spr/	6282	0.07
/ʃ/	4213	1.00	/st/	6252	0.08	/str/	5445	0.45
			/sk/	6260	0.08	/skr/	6107	0.15

Table 3.2: The model talker's naturally-produced centroid frequency (CF) and retraction ratio (RR) values for each phonological environment.

all onset instances of /s/ and /ʃ/ in the target words using a modified script originally created by DiCanio (2013). The average centroid frequency was calculated for each onset environment and the retraction ratio (see Section 2.1 for details) was subsequently calculated independently for each of the consonant clusters. The model talker's naturally-produced retraction ratio values are provided in Table 3.2. Recall that a value closer to 0 represents a more /s/-like sibilant while a value closer to 1 represents a more /ʃ/-like sibilant.

The retraction ratios presented in Table 3.2 suggest that the model talker produces a significantly more retracted /s/ in /str/ clusters, with a retraction ratio of 0.45, than /st/ clusters, with a retraction ratio of 0.08. This suggests that the model talker's /str/ is generally about halfway between his /s/ and /ʃ/, while /st/ is similar to prevocalic his /s/. Both /spr/ nor /skr/ clusters show minimal coarticulatory influences of /r/, with values similar to their corresponding /sC/ clusters. These observations serve to inform the creation of the manipulated auditory stimuli for this and the following experiments.

To create the stimuli for the increased and decreased retraction conditions in this experiment, the onsets for all /sC/ and /sCr/ clusters were manipulated to contain specified degrees of retraction. This was done by digitally mixing the naturally-produced onsets from prevocalic /s/ (*sip*) and /ʃ/ (*ship*) at different scaling ratios, using a Praat script originally created by Darwin (2005). For /str/ clusters: in the decreased retraction condition, the onset sibilant was comprised of 30% onset /ʃ/, a clear /s/ for my ears; in the increased retraction condition, the onset sibilant was 60% /ʃ/, at the /s/-/ʃ/ category boundary; and in the hyper-increased retraction condition it contained 90% /ʃ/, tipping it over my perceptual

	Decreased retraction		Increased retraction		Hyper-increased retraction	
	/s/	/ʃ/	/s/	/ʃ/	/s/	/ʃ/
/spr/	0.90	0.10	0.60	0.40	0.30	0.70
/str/	0.70	0.30	0.40	0.60	0.10	0.90
/skr/	0.90	0.10	0.60	0.40	0.30	0.70
			Across conditions			
			/s/	/ʃ/		
		/s/	1.00	0.00		
		/sp/	0.90	0.10		
		/st/	0.90	0.10		
		/sk/	0.90	0.10		
		/ʃ/	0.00	1.00		

Table 3.3: Scaling factors used in sibilant stimuli creation

boundary to /ʃ/. For both /spr/ and /skr/ clusters: in the decreased retraction condition, the onset sibilant was comprised of 10% /ʃ/, which I perceived as a clear /s/; in the increased retraction condition it was comprised of 40% /ʃ/, which was perceptually unusual without being consistently /ʃ/; and in the hyper-increased retraction condition, it was comprised of 70% /ʃ/, which was a noticeable /ʃ/. The /sC/ onsets did not differ between between conditions, and contained 10% /ʃ/ in the decreased, increased, and hyper-increased retraction conditions. The onset sibilants in prevocalic stimuli were cross-spliced, but not mixed, to reduce the potential effect of stimuli manipulation. The scaling factors used for the creation of each cluster can be seen in Table 3.3.

Stop-initial stimuli Since all the target stimuli are necessarily sibilant-initial, stop-initial stimuli were added to serve as a fillers to mask the nature of the experiment from the participants. Stop-initial stimuli were selected in order to allow for paradigms that mirrored the target stimuli, utilizing the same places of articulation as the intermediate stop in /sCr/ clusters and contrasting for the presence of absence of a subsequent /r/ while adding a voicing contrast in both the prevocalic and prerhotic environments.

This set-up yields a four-item filler paradigm for each place of articulation, e.g. for

	word	IPA	$SUBTL_{WF}$	lg10WF
/p/	pig	/pɪg/	39.14	3.30
/pr/	prick	/prɪk/	14.12	2.86
/b/	big	/bɪg/	682.82	4.54
/br/	brick	/brɪk/	10.18	2.72
/t/	tip	/tɪp/	27.63	3.15
/tr/	trip	/trɪp/	82.39	3.62
/d/	dip	/dɪp/	7.96	2.61
/dr/	drip	/drɪp/	5.12	2.42
/k/	kit	/kɪt/	17.65	2.95
/kr/	crypt	/krɪpt/	1.37	1.85
/g/	gift	/gɪft/	64.51	3.52
/gr/	grip	/grɪp/	9.69	2.69

Table 3.4: Stop-initial wordlist for Experiment I: Cue Integration

bilabial: *pig*, *prick*, *big*, and *brick*. The complete list of stop-initial stimuli is included in Table 3.4, grouped by place of articulation. Phonemic representations assume that alveolar obstruents preceding /r/ are phonemically stops, such that *drink* is phonemically /drɪŋk/ and not /dʒrɪŋk/ (see Section 2.3 for more on affrication in these environments).

There were two conditions for stop-initial stimuli: increased and decreased VOT. This assure that the fillers were also digitally manipulated like the target items. All voiceless stop-initial stimuli, both prevocalic and preceding /r/, were manipulated to contain the stipulated durations of aspiration, while the VOT remained constant for voiced stops in both conditions. Prior to manipulation, I examined the VOT for all naturally-produced target word-initial stops, voiced and voiceless, recorded by the speaker. The mean VOT

Voice onset time (s)							
/p/	0.071	/pr/	0.102	/b/	0.013	/br/	0.016
/t/	0.090	/tr/	0.117	/d/	0.024	/dr/	0.050
/k/	0.099	/kr/	0.120	/g/	0.034	/gr/	0.054

Table 3.5: Mean naturally-produced voice onset time values for the model talker

Voice onset time (s)					
	Decreased	Increased		Decreased	Increased
/p/	0.035	0.105	/pr/	0.050	0.155
/t/	0.045	0.135	/tr/	0.060	0.175
/k/	0.050	0.150	/kr/	0.060	0.180
Across conditions			Across conditions		
/b/	0.015		/br/	0.015	
/d/	0.025		/dr/	0.050	
/g/	0.035		/gr/	0.055	

Table 3.6: Voice onset time for increased and decreased VOT conditions

measurements for the naturally-produced word-initial stops are provided in Table 3.5. The model talker was relatively stable in his VOT production, with few environments showing significant variation. The sole exception was the production on /dr/ clusters, which showed a high degree of variation with VOT values ranging from 0.03 to 0.84 ms. This variation was also perceptible auditorily, with some instances perceived as containing an affricate /$\widehat{\text{dʒ}}$/ and other perceived as containing a stop /d/.

The voiceless stimuli onsets were manipulated to contain an increased or decreased VOT, with aspiration duration respectively at 0.5 and 1.5 times the model talker's naturally-produced mean duration for that environment. In the increased VOT condition, VOT was extended by cross-splicing stable medial periods of aspiration from a different token into randomly selected time points in the medial portion of the aspiration of the target waveform, following Shockley et al. (2004) and Yu et al. (2013). In the decreased VOT condition, VOT was reduced by randomly removing medial periods of aspiration from the target waveform. This methodology yielded the VOT for the stimuli provided in Table 3.6.

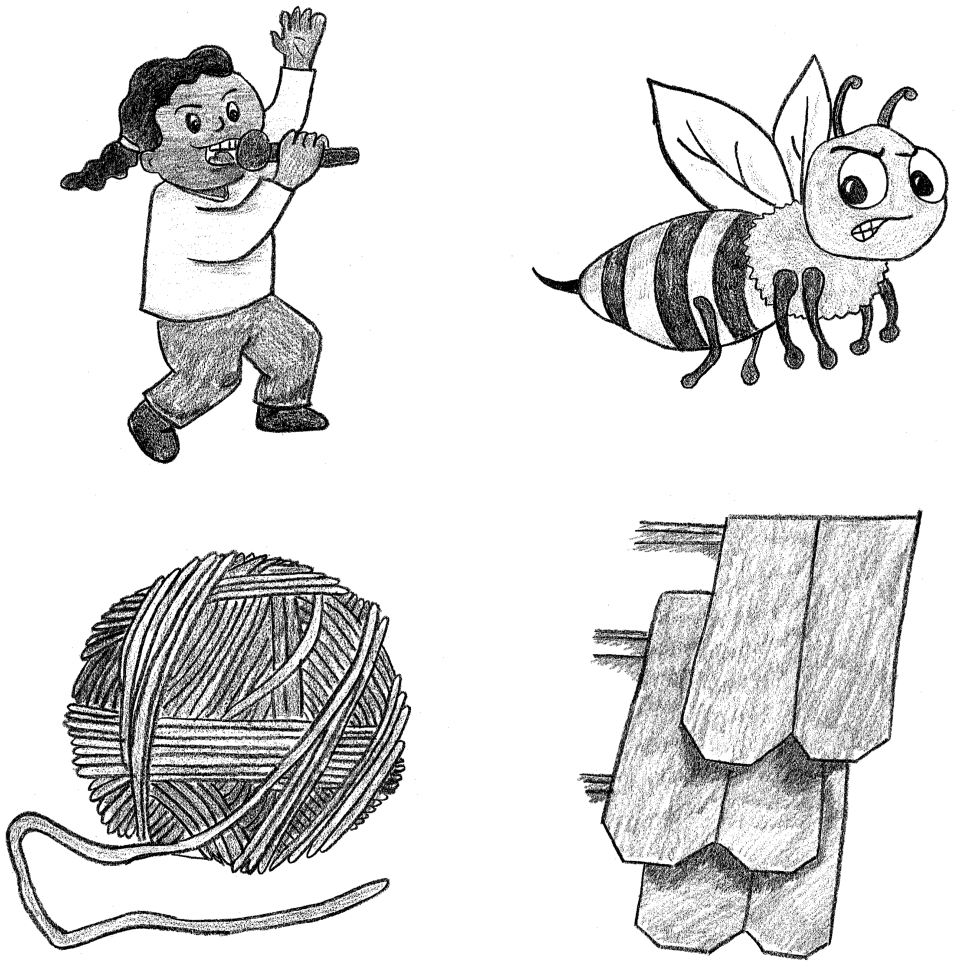


Figure 3.1: Sample visual stimuli for 'sing' (top left), 'sting' (top right), 'string' (bottom left), and 'shingle' (bottom right). All images are provided in Appendix A.2.

Visual stimuli Each auditory stimulus was also paired with a corresponding visual stimulus to represent the lexical item in the Visual World Paradigm. For each target word, I selected three free and publicly available clipart-style images. The images were then resized and gray-scaled. Four naïve non-linguist volunteers selected the image that best corresponded the intended word. Most images were selected unanimously, but in the case of a tie, I cast the tie-breaking vote. In order to control for differences of style, darkness, or image resolution, I then drew all images by hand using grayscale colored pencils, making adjustments to remove any text or distracting features. The hand-drawn images were then

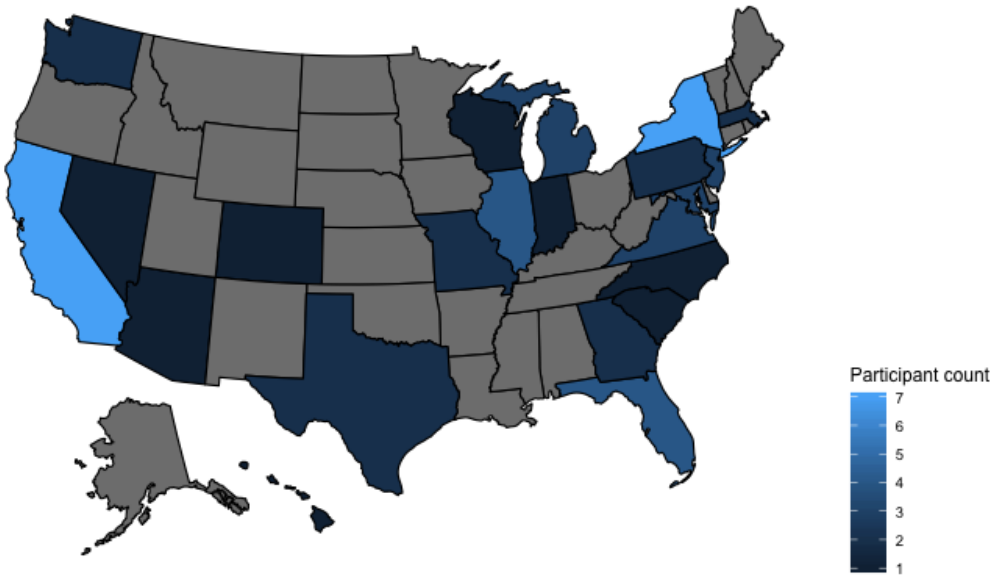


Figure 3.2: Geographic distribution of participants by self-reported state lived in the longest before age 18, with a lighter blue indicating a higher count of participants. Gray indicates no participants reporting living in that state.

scanned and resized to 550x550 pixels. A sample paradigm for /str/ is presented in Figure 3.1 and all images are provided in Appendix A.2.

3.3.2 Participants & procedure

Participants

Fifty-two participants were recruited from the University of Chicago and the greater Chicago area. Initial participants were randomly assigned to either the increased or decreased retraction condition, as well as the increased or decreased VOT condition. Later recruits were automatically assigned to the hyper-increased retraction condition and randomly assigned to either increased or decreased VOT condition. This yields 6 possible combinations of retraction and VOT conditions.

All participants were between the age of 18 and 22 (mean=20). Undergraduate students enrolled in introductory linguistics classes participated for course credit. Individuals not par-

ticipating for credit were paid at the rate of $20 per hour. All participants were self-reported native speakers of American English and grew up predominately in the United States. The geographic distribution of the participants is mapped in Figure 3.2, with notable gaps in much of New England, the plains, the Rockies, and the inland South. Most participants self-identified as growing up in a suburban area (34), with fewer from urban (15) or rural (3) environments. More participants identified as female (37) than male (15), and no participants identified as non-binary or transgender. Just more than half of the participants identified as straight/heterosexual (29) and/or white (29). Additional demographic information concerning a variety of social and cognitive metrics, obtained through a series of surveys, is included later in this section.

No participants included in the analysis reported a history of hearing loss, language and communication disorders, stroke, traumatic brain injury, other medical or neurological conditions commonly associated with cognitive impairment. An additional nine individuals participated in the present experiment but were excluded due to non-native status, reported language or neurological disorders, and/or non-attentive responses.

Procedure

The experiment took place in the Language Processing Laboratory at the University of Chicago, in a quiet room equipped with a Tobii T-60 eye-tracker.

Prior to beginning the lexical identification task, participants were first familiarized with the images and their associated lexical items, as it is necessary to present the stimuli during the lexical identification task without any accompanying orthographic representation. While the nouns were generally easier to represent visually, the verbs and adjectives can be more challenging, like *big*, which was represented by an image of two elephants. For other items, it was necessary to train participants to associate the image with a lower frequency word that fits the paradigm rather than a more obvious high frequency word, for example listeners

	s–ʃ	s– sC	sC–sCr	sCr–ʃ
/p/	sit–shit	sit–spit	spit–spritz	spritz–shit
/t/	sing–shingle	sing–sting	sting–string	string–shingle
/k/	sip–ship	sip–skip	skip–script	script–ship

	T–D	T–Tr	Tr–Dr	D–Dr
/p/	pig–big	pick–prick	prick–brick	big–brick
/t/	tip–dip	tip–trip	trip–drip	dip–drip
/k/	kit–gift	kit–crypt	crypt–grip	gift–grip

Table 3.7: Pairing of sibilant- and stop-initial stimuli

must associate a picture with *crypt*, which is a filler item, rather than *tomb*. All images are provided in Appendix A.2.

Following Beddor et al. (2013), participants were first introduced to the images with their accompanying labels. Participants were ask to read the label aloud and explain to the researcher the relationship between the image and the label. As an example, for picture of a dog with the label *dog*, the researcher would say "That's a dog", and for a picture of a cheetah with a label *fast*, they would say "Cheetahs are fast". Following this first task, participants were shown the images without their accompanying labels and were asked to reproduce the corresponding label. All participants had 100% accuracy in the label reproduction task, demonstrating that they had successfully associated the lexical words and images.

For the identification task, participants were seated in front of a Tobii T-60 eye-tracker, with a sampling rate of 60Hz. The eye-tracker was recalibrated for each participant. Participants viewed images on the screen in a modified Visual World Paradigm (Allopenna et al., 1998), containing two images rather than the more typical four, following Beddor et al. (2013). The images were paired according to the contrasts designed and explained in Tables 3.7, for example *sting* vs. *string*. Participants were given two seconds to scan the screen to identify the images. After scanning the images, participants were required to focus on a fixation cross in the center of the screen, equidistant between both images. Once their fixation on the cross was confirmed, a red box was displayed. Participants then clicked

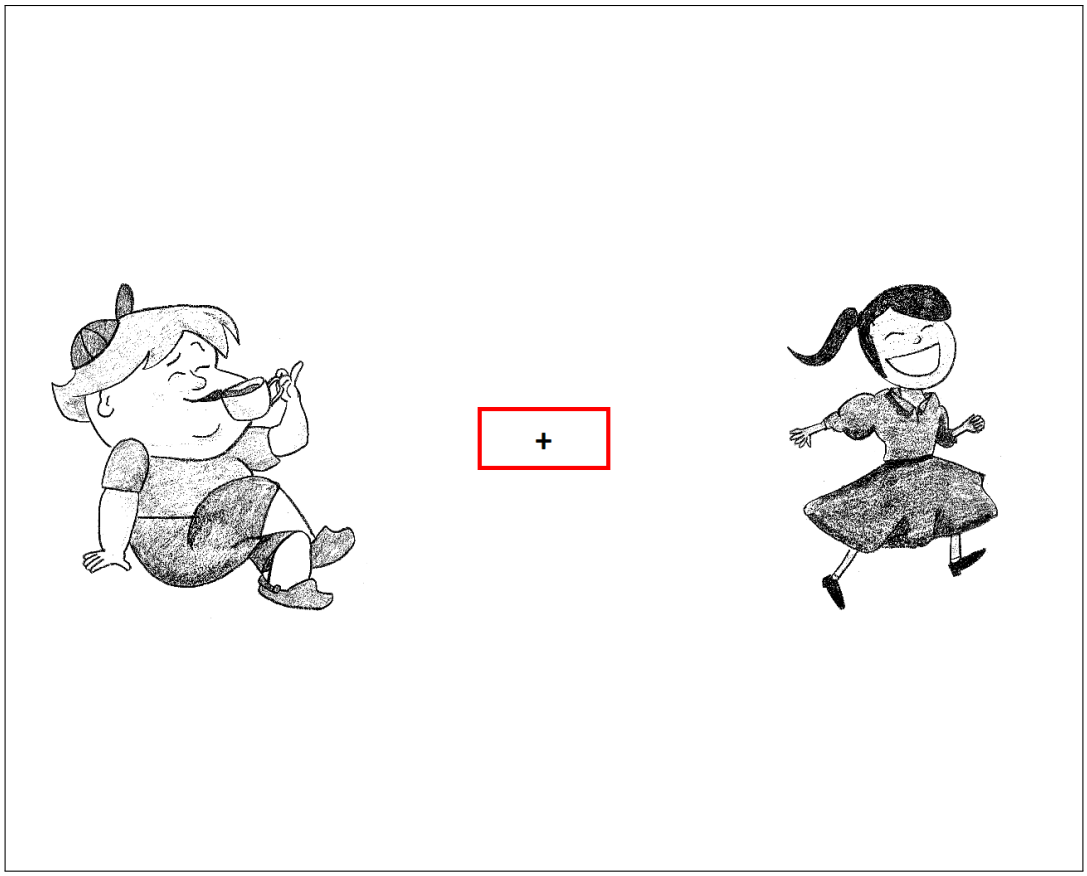


Figure 3.3: A sample trial illustrating the presentation of the images, *sip* (left) and *skip* (right) and fixation cross, with the subject's gaze detected on the fixation cross, indicated by the red box.

within the box to receive auditory instructions in the form "Now select [word]", for example "Now select *sting*". Participants were then directed to click on the corresponding image as quickly as they could, with their mouse and eye movements recorded throughout. The trial ended when either image was selected and the experiment automatically advanced to the next item. Each trial lasted roughly 5 seconds. A sample trial slide is provided in Figure 3.3 to illustrate how the visual stimuli and response options were presented.

Participants completed five practice trials with the researcher present before continuing onto the experimental trials when the researcher left the room. There were 96 trials per block with two repetitions of each pair in Table 3.7, with each item of the pair presented as the target. There were four blocks in total with breaks between the blocks determined

by the participant. This yields 384 trials in total and an average total time of around 32 minutes for the task.

Demographic survey

After the experiment proper, participants completed an extensive demographic survey. This included the questions used to determine eligibility, like native speaker status and history of language/neurological disorders or impairments. This survey was also used to determine participants' self-reported demographic categories, like age, gender, sexuality, race, and place of origin. For each of these demographic questions, response options were presented with radio buttons, including an option for a self-described alternative response. A summary of these demographic responses was provided at the beginning of this section. The complete post-test survey is provided in Appendix A.4.

Following these more standard demographic categories, participants completed a series of sub-surveys to determine their relative measurements on a variety of social and cognitive scales. These included the Male Role Attitude Scale (MRAS, Pleck et al., 1993), Empathy Quotient (Baron-Cohen & Wheelwright, 2004), Big Five Inventory (John et al., 2008), and PROMIS Anxiety score (Cella et al., 2010). Each of these surveys are provided in Appendix A.4. These different measurements were selected in order to account for variation in speech perception strategies. Especially given the between-subject design of this experiment, it is crucial to account for potential variation between the condition groups that may explain differences in cue integration strategies.

Firstly, the Male Role Attitude Scale (MRAS: Pleck et al., 1993) is a standardized psychological instrument designed to to discern an individual's relative endorsement of various stereotypes of masculinity. Following the results of the online meta-commentary examination conducted in Chapter 2.6, this survey is included to account for potential influences in speech processing based on an individual's relative endorsement of traditonal stereotypes of

toughness. In particular, I focus on the toughness subscale, as defined by two questions: *A guy will lose respect if he talks about his problems* and *A young man should be physically tough, even if he's not big.* The MRAS and the toughness subscale specifically have been used as explanatory variables in the social evaluation (Levon, 2014) and phoneme categorization (Campbell-Kibler, forthcoming) of prevocalic sibilants. The entire survey is provided in Appendix A.4. Campbell-Kibler (forthcoming) found that individuals who more strongly endorse masculine stereotypes account for perceived performances of masculinity in their perception of onset sibilants more than listeners who reject such stereotypes. Thus, a listener who endorses masculine stereotypes is more likely to categorize the same ambiguous sibilant as /s/ when the speaker is perceived as more masculine and /ʃ/ when the speaker is perceived as less masculine. It is possible that individuals who more strongly endorse such stereotypes may be more likely to attribute coarticulatory information to social performances and thus unable to use those cues to determine and predict upcoming phonological content. The mean composite score for the present experiment was 14.98 (s.d. 3.75, on a scale of 10 to 40) and a mean score of 2.78 (s.d. 1.19, on a scale of 2 to 8) on the toughness subscale, with a higher score indicating a stronger endorsement of the relevant stereotypes. This suggest on the whole whole, student participants at the University of Chicago are much more likely to reject traditional stereotypes of masculinity, including masculine stereotypes of toughness.

The Empathy Quotient (EQ: Baron-Cohen & Wheelwright, 2004) is a short, 22 question self-administered survey for identifying the degree to which individuals of normal intelligence exhibit traits commonly associated with Autism Spectrum Disorder (ASD). Specifically, the EQ examines the degree to which respondents are able to identify and respond to another individual's thoughts and emotions. The EQ was not used to diagnose ASD. It was solely employed to represent a participant's relative identification with traits commonly associated with ASD. The EQ, and the more extensive Autism Quotient (which includes traits such as attention to detail, attention-switching, communication, etc.), have previously been illus-

trated to influence individuals' speech perception strategies, and specifically their likelihood of compensating for coarticulation (Stewart & Ota, 2008; Yu, 2010, 2013). Yu (2013) finds that individuals with a lower EQ, i.e. individuals who empathize less, are less likely to account for phonological context in their sibilant categorization. It is possible that individuals with a lower EQ who are less able to compensate for coarticulation are also less able to use context-dependent cues to anticipate upcoming sounds. The mean score for the present experiment was 26.05 (s.d. 8.69), on a scale of 0 to 44, with a higher score indicating that individual is more empathetic.

The Big Five Inventory (John et al., 2008) assesses personality traits along five dimensions: Extroversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. Extroversion is characterized by positivity and the desire to seek out the company of others; typical facets associated with extroversion are sociability, assertiveness, gregariousness, and social confidence. Agreeableness is characterized by a compassionate and cooperative approach to working with others; typical facets associated with agreeableness are: trust, warmth, generosity, and modesty. Conscientiousness is characterized by the tendency to aim for achievement goals; typical facets associated with conscientiousness are orderliness, decisiveness, industriousness, and self-discipline. Neuroticism is characterized by the tendency to exhibit emotional instability; typical facets associated with neuroticism are anxiety, irritability, depression, and hostility. Finally, openness is characterized by an appreciation of art, adventure, imagination, and the unknown or unfamiliar; typical facets associated with openness are intellectualism, idealism, adventurousness and imagination. The inclusion of the Big Five Inventory was primarily motivated for Experiment III: Convergence (Chapter 5), as Yu et al. (2013) found that individuals who have who score higher on the openness scale are more likely to exhibit phonetic convergence. It is possible that individuals who are more likely to converge are also more sensitive or receptive to their interlocutor's cues and thus better able to use those cues as soon as they become available. The mean openness score for

the present experiment was 28.82 (s.d. 5.92) on a scale from 15 to 40, with a higher score indicating that the subject is more original, inventive, imaginative, etc.

The PROMIS Anxiety instrument (Cella et al., 2010) is a question bank to assess self-reported anxiety, including fear (fearfulness, panic), anxious misery (worry, dread), hyper-arousal (tension, nervousness, restlessness), and somatic symptoms related to arousal (racing heart, dizziness). There are multiple questions banks calibrated for different populations, and the adult short form for anxiety–emotional distress was selected for the present study. The survey consisted of 29 questions. Assessments of anxiety were included in the demographic survey as previous research has found that heightened anxiety diminishes phoneme discrimination and categorization. Mattys et al. (2013) induced the physiological responses to anxiety by having participants inhale air enriched with 7.5% CO_2, finding that listeners exhibited similar decreased performance when anxiety was induced as when their attention was divided between multiple tasks. It is possible that individuals who are more anxious, and thus exhibit diminished phoneme discrimination and categorization, are less able to use context-dependent cues in real-time to aid in speech processing. The mean score for the present experiment was 64.07 (s.d. 18.82) on a scale from 29 to 145, with a higher score indicating that the subject reports experiencing more anxiety-related traits in the past 7 days.

The complete post-test survey took average of 10 minutes to complete. The entire session, including informed consent, instructions, experiment, and demographic survey lasted approximately 50 minutes.

3.3.3 Measurements, analyses, & hypotheses

Behavioral measurements

Due to the nature of the stimuli, no trial was expected to be ambiguous; regardless of the degree of retraction observed on the onset sibilant, the ultimate presence or absence of

/r/ would disambiguate the stimulus. Accuracy, defined as clicking on the correct image, for each target stimuli pairing in the different retraction conditions is presented in Figure 3.4. Visual inspection of the figure suggests that all pairings are approaching ceiling, with accuracy above 95% in all stimuli pairings and the confidence interval straddling 100% in all but two of 72 pairings. As all responses were at or approaching ceiling, behavioral data was not analyzed, and instead instances in which the incorrect image was selected were excluded from the gaze data analysis. However, it is worth noting that one of the two confidence intervals not straddling 100% is *sting* vs. *string* in the decreased retraction condition (mean = 0.958; 95% confidence interval = 0.922 – 0.996), precisely where listeners may most expect to use the coarticulatory cues of retraction but find them less available.

Gaze measurements

Participants' eye gaze was monitored from the initial display of the target and competitor images, through the cross fixation, until 2000 ms following the onset on the target word or until they clicked on one of the images to make a selection, whichever came first. Although eye gaze was tracked for both left and right eyes, analysis was conducted on the right eye exclusively. Unlike trial accuracy, which identifies whether the participant selected the correct image corresponding to the auditory stimuli, gaze measurements identify precisely when the target or competitor lexical item were considered, before the ultimate decision to click on the correct image was made. This not only provides a much more fine-grained temporal resolution than reaction time for mouse clicks, but also allows for an examination of alternative phonological candidates for the (ultimately) unambiguous stimuli.

The online measurement selected for analysis for the present experiment was the proportion of correct fixations over time, which is determined by examining the accuracy of each individual fixation. A fixation was determined to be a correct fixation if the right eye gaze fell within the 550x550 pixel region containing the image corresponding to the auditory

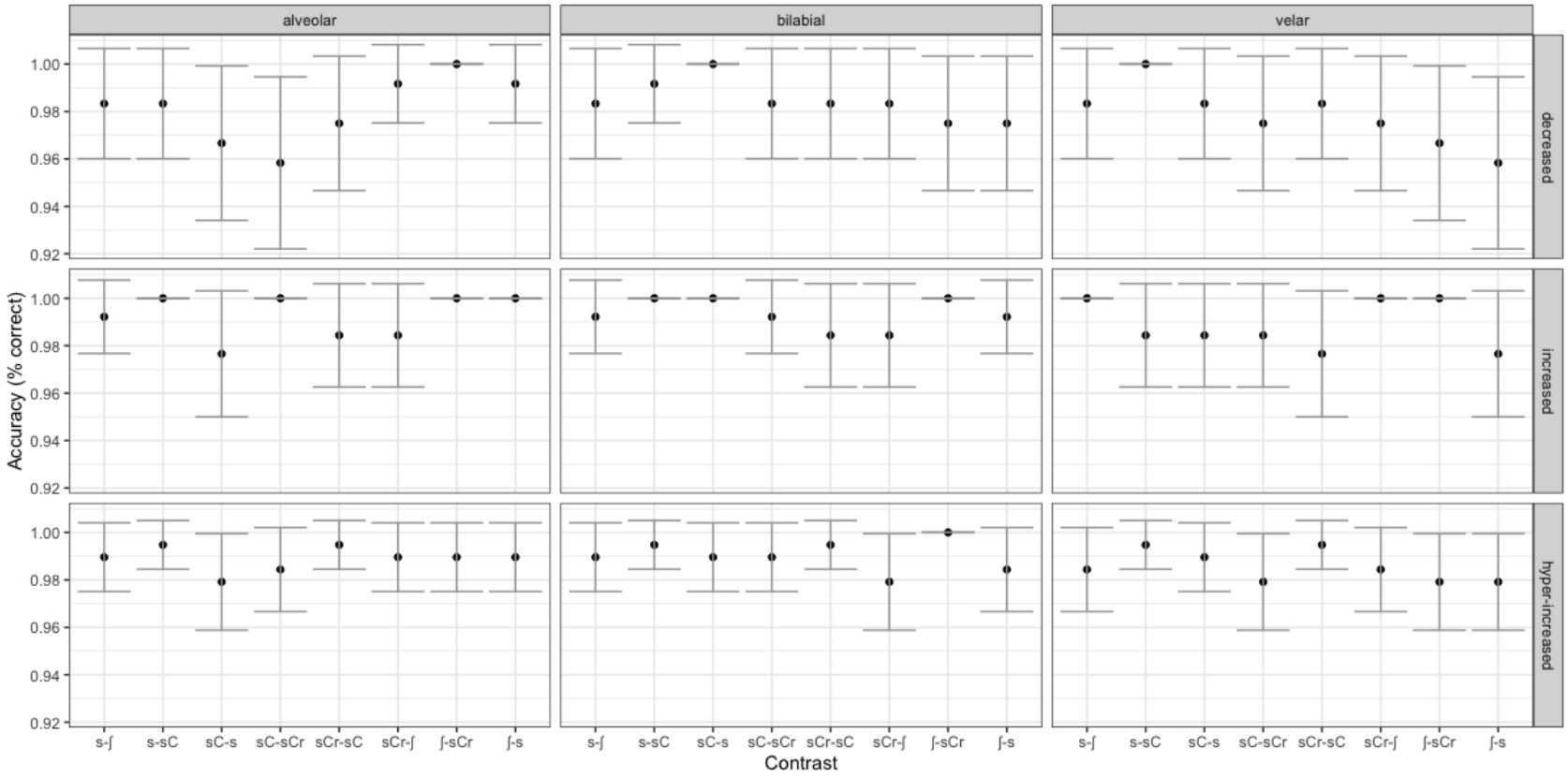


Figure 3.4: Accuracy (y-axis) for each stimuli pairing (x-axis) by retraction condition (panel row) and place of articulation (panel column). Note that the y-axis has been altered to highlight the minute differences observed between conditions (0.92 to 1.0), and all values are approaching ceiling.

stimuli. Fixations were binned into 20 ms windows. A proportion of 0 for a given bin means that there were no trials in the relevant condition during which eye gaze was detected within the 550x550 pixel region containing the target image. This means that all participants' gaze was directed at the fixation cross, the competitor image, or anywhere else on the screen other than the target image. Thus, it is not the proportion of target versus competitor fixations, but rather the proportion of target versus non-target fixations. Similarly, a proportion of 1 means that in all trials a target fixation was detected within the specified 20 ms window.

Analysis

Generalized linear mixed-effects models with a logit link function were fit to the accuracy of a given fixation (1,0) using the `glmer()` function in the `lme4` package (Bates et al., 2015) in `R` (R Core Team, 2015). Separate models were fit on the different onset pairings presented to participants, including: /s/–/ʃ/ and /sC/–/sCr/. It takes approximately 200 ms to plan and execute an eye movement and the total sibilant duration was 180 ms. Thus, the models were run on the 180 ms window that began 200 ms following the onset of the stimulus sibilant, representing any eye movements that were planned during the sibilant. As fixed effects, the prevocalic model fit on /s/–/ʃ/ pairings includes Onset (/s/ and /ʃ/; treatment-coded with /s/ as base), while the preconsonantal model fit on /sC/–/sCr/ includes Place of articulation (alveolar, bilabial, and velar; Helmert-coded with alveolar as base) and ClusterType (/sC/ and /sCr/; treatment-coded with /sC/ as base). Both models additionally include trial Order (1–384, scaled), TimeWindow of the sibilant (1–180, binned into 20 ms windows and scaled), and RetractionCondition (decreased, increased, and hyper-increased, treatment-coded with decreased as base; counterbalanced between subjects). Self-reported responses for the demographic categories Gender (male, female; sum-coded), Sexuality (straight, queer; sum-coded), and Region (midwest, northwest, south and west, treatment-coded with midwest as base) were included to capture potential social variation. Addition-

ally, each of the social and cognitive scores that are predicted to influence cue integration were included in each model including ToughnessEndorsement[2], Anxiety, Empathy, and Openness. Each score was scaled, an in the case of Anxiety inverted, such that a higher value corresponds with a higher predicted phoneme discrimination.

Preliminary models for the different onset pairings included all two-, three-, and four-way interactions between the fixed effects predictors. However, the models failed to converge and the analysis was thus limited to two- and three-way interactions. All interactions that did not reach a significance threshold of 0.05 were pruned from the final model. Additionally, the preliminary models included maximally specified random effects structures, with by-subject random slopes and intercepts, which were progressively simplified until convergence was achieved. Final models for the different onset pairing are reported in the following section.

Hypotheses

The specific hypotheses for participants' eye gaze are as follows:

Hypothesis 1 The first hypothesis predicts that listeners make immediate use of spectral cues to distinguish /s/ and /ʃ/ in prevocalic environments. This hypothesis directly tests the buffer strategy observed by Galle et al. (2019). Under this hypothesis, correct fixations on /s/ or /ʃ/ will emerge during the sibilant interval, before formant transition information becomes available. This is directly tested by TimeWindow in Model 3.1, as if listeners improve their proportion of correct fixations over the sibilant interval, it suggests that they are using the cues available to them from the sibilant. If a cascade strategy is observed for prevocalic sibilants, this suggests that listeners can use the spectral cues of sibilants as soon as they are available, contra Galle et al. If a buffer strategy is used, then we need

2. Separate models were fit on both the composite MRAS score and the toughness subscale, with the toughness models outperforming models fit on the composite score. This suggests that endorsement and expectations for masculine toughness influence speech processing of onset sibilants more so than masculine stereotypes in general.

to turn to preconsonantal environments (Hypothesis 2-4) to ask if a cascade strategy can be observed in those environments. If a buffer strategy is used in both preconsonantal and prevocalic environments, each of Galle et al.'s potential explanations remain possible. And finally, if a cascade strategy is used in preconsonantal but not prevocalic environments, this challenges the auditory grouping account, i.e. sibilants are acoustically different from other speech sounds, as preconsonantal sounds are acoustically within the range of the prevocalic sibilants. Furthermore, this challenges the vowel-as-organizer account, as listeners do not need to wait for the vowel when it's sufficiently distant.

Hypothesis 2 The second hypothesis proposes that listeners use coarticulatory cues in predicting the phonological context of a sibilant, and do so as soon as those cues are available. A confirmation of this hypothesis would demonstrate that listeners use a cascade (immediate integration) strategy of speech processing for long distance sibilant-rhotic coarticulation. This would mirror the cascade strategy observed for vowel-nasal coarticulation (Beddor et al., 2013). Under this hypothesis, correct fixations to either the /sCr/ or /sC/ will emerge during the sibilant interval, before the /r/ is perceived. Like in the prevocalic model, we again test this by looking to TimeWindow but this time for the preconsonantal model in 3.2. Furthermore, we look to the possible effect of RetractionCondition (decreased, increased, and hyper-increased) in conditioning such fixations, as the greater the degree of retraction, the stronger the cues of the upcoming /r/. If Hypothesis 2 is confirmed, then the following hypotheses stand to be tested:

Hypothesis 3 The third hypothesis predicts that a retracted /s/ is a better indicator of rhotic presence than a non-retracted /s/ is for rhotic absence. That is, does a more retracted, i.e. more /ʃ/-like, onset predict an /sCr/ cluster better than a less retracted, i.e. more /s/-like, onset predicts an /sC/ cluster. At its core this proposes that listeners not only use context-dependent knowledge about sibilant variation (see Hypothesis 2), but that

/s/-retraction specifically is useful to and used by listeners. A confirmation of this hypothesis would demonstrate that the cues of /s/-retraction are more useful in speech processing than the absence of such cues, much like the findings of Beddor et al. (2013) that a nasal vowel is a better cue of an upcoming nasal stop than an oral vowel is of an upcoming oral stop. Under this hypothesis, correct fixations to /sCr/ clusters should be faster than correct fixations to /sC/ clusters. This is tested directly by ClusterType in Model 3.2.

Hypothesis 4 The fourth and final hypothesis predicts that the cues of /s/-retraction are a better indicator of rhotic presence in /str/ clusters compared to /spr/ and /skr/ clusters. A confirmation of this hypothesis would demonstrate that listeners have detailed phonological knowledge about /s/-retraction as a sound change in progress, with greater degrees of retraction observed in /str/ clusters (Baker et al., 2011), and adjust their expectations accordingly. To test this, we look to the an effect of Place of articulation (alveolar, bilabial, and velar) in conditioning looks toward the /sCr/ image during the sibilant interval. If this hypothesis is confirmed, listeners will look to /str/ clusters faster than /spr/ and /skr/ clusters. Crucially, this hypothesis assumes that listeners use their contextual knowledge about the upcoming stops, in this task provided through the two images displayed, in order to use stop-specific strategies before the stop itself has been perceived.

3.4 Results

The results of this experiment are presented in two parts. First, in Section 3.4.1, I present the results of the prevocalic /s/-/ʃ/ pairings, which speak to the first hypothesis, and ask whether listeners can use the spectral cues to distinguish /s/ and /ʃ/ immediately or whether they use a buffer strategy. Next, in Section 3.4.2, I present the results from the /sCr/-/sC/ pairings, which speak to the final three hypotheses, and ask whether listeners can anticipate the presence of the upcoming /r/ in real time. In addition to the group results, I briefly

examine individual variation in speech processing patterns for these cluster onsets.

3.4.1 Prevocalic results

Before delving into the context-dependent variability and complexity of the preconsonantal environments, the first model asks if listeners can immediately use spectral cues when considering /s/, e.g. *sip*, and /ʃ/, e.g. *ship*, candidates. The first model presented is fit on the accuracy of individual fixations planned during the 180 ms sibilant interval for prevocalic /s/ and /ʃ/. Unlike the preconsonantal model, the present model has just one predictor for the Onset category (/s/ or /ʃ/). The final model for correct fixations selected after progressive simplification of the random effects structure in order to achieve model convergence is presented in `lme4` format in Formula 3.1.

$$\begin{aligned}&\textsc{Fixation} \sim \textsc{Order} + (\textsc{TimeWindow}+\textsc{Onset}+\textsc{RetractionCondition})^3+\\&\textsc{Region}+\textsc{Gender}+\textsc{Sexuality}+\textsc{Tough}+\textsc{Empathy}+\textsc{Openness}+\textsc{Anxiety}+\\&(1+\textsc{Order}+\textsc{Onset}|\textsc{Subject})\end{aligned} \tag{3.1}$$

This model asks if listeners are more likely to look at the correct image during the sibilant interval, given the Onset phonemic category, the RetractionCondition, and the 20 ms TimeWindow examined. The social and cognitive predictors are included to account for any potential individual variability, especially given the between-subject design of the present experiment, but are not the primary interest of the experiment. Experiment II: Categorization (Chapter 4) seeks to fill that gap, asking what potential socio-indexical meaning /s/-retraction bears. The inclusion of by-subject random intercepts and by-subject random slopes for trial Order and Onset suggest significant individual variability with respect to these predictors. By-item random slopes and intercepts are not included as there is only one item per onset cluster, given the training and time constraints of the current design. The

Table 3.8: Model predictions for all significant mains effects and interactions in fixation accuracy for /s/ vs. /ʃ/ onsets, N=26205. A positive value indicates a greater prediction of fixations on the target word. Complete model predictions including variables and interactions that did not reach a significance threshold of 0.05 are included in the Appendix as Table A.1.

	Est.	*SE*	*t*	*p*
Order	0.05	0.01	3.45	< 0.001
TimeWindow	0.27	0.04	7.31	< 0.001
TimeWindow:Increased	0.19	0.05	3.79	< 0.001
TimeWindow:Hyper	0.12	0.05	2.63	0.002

significant effects and interactions of the fixation accuracy model for prevocalic sibilants are presented in Table 3.8.

Unlike the preconsonantal model, a significant effect of trial ORDER is observed ($t = 3.45, p < 0.001$), suggesting that individuals are more likely to look at the correct image as the experiment progresses. This may indicate that as the listener becomes familiarized with the model talker's speech, they are better able to quickly make decisions about the intended target using spectral cues alone. Furthermore, there is no significant effect of ONSET phoneme category, suggesting that listeners are equally likely to direct their eye gaze toward the target word when the onset is /s/ as when it is /ʃ/. However, as the by-subject random slopes for ONSET improved the model likelihood, this hints to individual variation with respect to possible differences in the processing of the different sibilants.

Although the model is fit on the accuracy of the individual fixations, for the sake of data visualization, I will present the proportion of /ʃ/ fixations. I do so primarily because this allows for the visual representation of the different onsets to diverge precisely when the consideration of the candidates themselves diverge. Plotting the proportion of correct fixations instead would be characterized by overlapping points throughout the stimulus interval. Recall that the prevocalic sibilants were cross-spliced but not manipulated, so that a prevocalic /s/ is always a naturally-produced prevocalic /s/ and a prevocalic /ʃ/ is always a naturally-produced prevocalic /ʃ/. Thus, unlike in the preconsonantal environment, /s/ and

/ʃ/ onsets are potentially immediately disambiguated, as has been shown in gating tasks (Galle et al., 2019, i.a.), but it remains to be demonstrated that listeners can immediately disambiguate them in speech processing.

Figure 3.5 asks how listeners' gaze to either the /s/ or /ʃ/ onset changes the course of the trial. Time after the sibilant onset is indicated on the x-axis and the proportion of looks at the /ʃ/ image is indicated on the y-axis. A trial with an /s/ onset is indicated in teal and a trial with an /ʃ/ onset is indicated in red. The three panels are present the counterbalanced retraction conditions, although prevocalic onsets were not manipulated between retraction conditions. For both /s/ and /ʃ/ onsets, listeners begin on even footing, with one quarter of fixations to /ʃ/ onsets. The remaining fixations are not indicated in Figure 3.5, but again about one quarter of fixations are to /s/ onsets, while half of fixations are on neither image. Remember that to initiate the trial, participants are required to fixate on a cross midway between both images, which may explain why the plurality of looks are on neither image at the sibilant onset. Thus in Figure 3.5, a correct fixation for /ʃ/ onsets is indicated directly by a higher proportion of /ʃ/ fixations, while a correct fixation for /s/ onsets may be indicated indirectly by a decreased proportion of /ʃ/ fixation.

The fixation proportion is provided from the sibilant onset to one second after that onset. Given that it takes approximately 200 ms to plan and execute an eye movement, fixations planned during the sibilant interval would be observed approximately 200 ms later. Vertical lines are provided to serve as guideposts for what sound was heard when the eye movement was planned. If a look to an /ʃ/ cluster was planned during the sibilant, it would be indicated in Figure 3.5 by an increased proportion of /ʃ/ fixations between the dashed lines, which represent a 200 ms delay from the sibilant onset and vowel onset respectively. A look once the vowel has been heard and formant transition information becomes available is indicated by an increased proportion of /ʃ/ fixations after the second dashed line.

Preliminary inspection of Figure 3.5 may first highlight that proportion fixations never

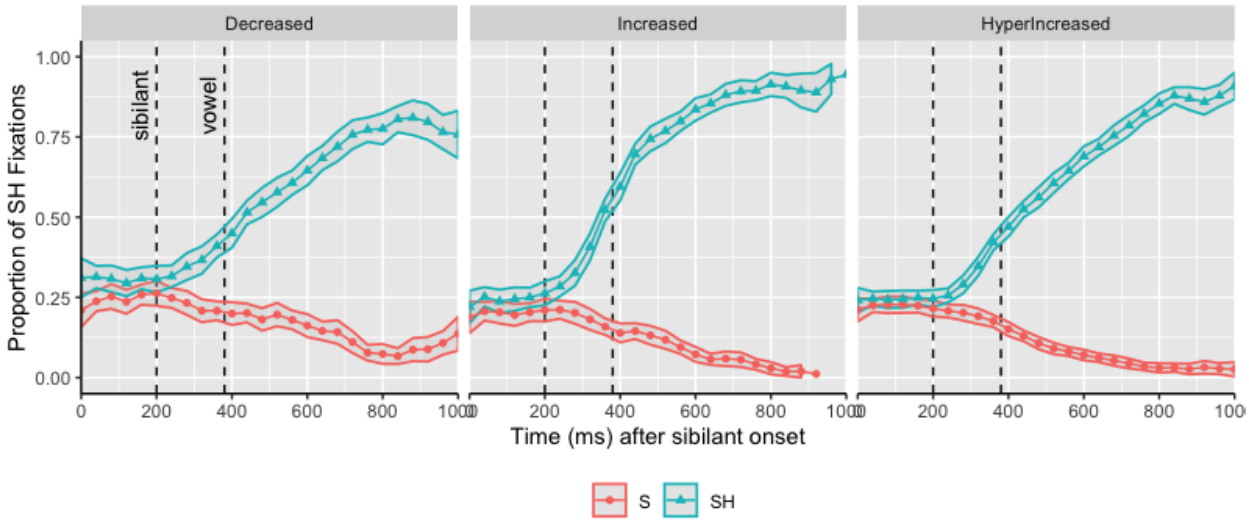


Figure 3.5: Fixation proportion for /ʃ/ clusters (y-axis) by time following the sibilant onset (x-axis, binned into 40 ms windows) and onset category (color: /s/ = red circles, /ʃ/ = teal triangles) and retraction condition (panel row). The vertical lines represent a 200 ms delay from the onset of the sibilant and vowel. A fixation executed during the sibilant interval would be observed between the black dashed vertical lines.

reach 1, that is, there is no 20 ms window in which all participants are looking at the target item, despite the behavior results at ceiling. This is because as one participant identifies and selects the target item, another participant may not yet have made that decision, while another still has already done so and directed their gaze elsewhere. Furthermore, preliminary inspection may focus on the observation that the proportion of correct fixations does not rise dramatically until well after the vowel onset. This is certainly the case, with many trials not exhibiting eye gaze on the target item until after the sibilant interval, which suggests that, in many instances, listeners wait until formant transitions are available to identify the target word, consistent with the findings of Galle et al. (2019). However, the present experiment is concerned with the relative accuracy *during* the sibilant interval. I ask, can listeners use the spectral cues available during the sibilant interval to contrast /s/ and /ʃ/? At the most fundamental level, this examines whether accuracy of fixations increases over the course of the sibilant. To see this, we look for diverging predictions and steep slopes for /s/ and /ʃ/ between the dashed lines.

Visual inspection of Figure 3.5 first show clear positive slopes for /ʃ/ trials, in teal, and clear negative slopes for /s/ trials, in red, across all RetractionConditions. Additionally, the confidence intervals for /ʃ/ and /s/ visibly diverge in all three RetractionConditions, and do so almost immediately during the sibilant interval. This is supported by the model, with a robust effect of TimeWindow ($t = 7.31, p < 0.001$), which suggests that the proportion of correct fixations increases over the duration of the sibilant. Further visual inspection Figure 3.5 indicates that the relative increase in /ʃ/ fixations over the course of the sibilant is greatest in the increased retraction condition in the center panel, and slightly more pronounced in the hyper-increased retraction condition compared to the decreased retraction condition. This is confirmed by the model, with a significant interaction of TimeWindow with RetractionCondition (increased: $t = 3.79, p < 0.001$; hyper-increased: $t = 2.63, p = 0.002$). Although the prevocalic sibilants themselves are not manipulated and do not differ between the different retraction conditions, these findings suggest that listeners are actually faster at distinguishing prevocalic /s/ and /ʃ/ when the model talker exhibits more dramatic context-dependent sibilant variation in other phonological environments. This finding, while unexpected, may suggest that listeners find spectral cues for sibilants more reliable and useful for speakers who produce more systematic and predictable context-dependent variation than for listeners who do not vary at all.

3.4.2 Preconsonantal results

With the results of the first model confirming that individuals can immediately use the spectral cues of sibilants when making phonemic decisions between /s/ and /ʃ/, the second model asks if they can use the context-dependent cues of sibilant variation to predict the phonological environment of the sibilant. The second model is fit on the accuracy of individual fixations during the 180 ms sibilant interval for /sCr/ and /sC/ clusters. This model asks whether listeners immediately use spectral cues when considering /sCr/, e.g. *string*,

and /sC/, e.g. *sting*, candidates. This model thus separates onset clusters into two different predictors: PLACE of articulation (alveolar, velar, or bilabial) and CLUSTERTYPE (/sC/ or /sCr/). The final model for portion of correct fixations selected after progressive simplification of the random effects structure in order to achieve model convergence is presented in `lme4` format in Formula 3.2.

$$\begin{aligned}&\text{FIXATION} \sim \text{ORDER} + (\text{TIMEWINDOW}+\text{PLACE}+\text{CLUSTERTYPE}+\text{RETRACTIONCONDITION})^3 \\ &+ \text{REGION}+\text{GENDER}+\text{SEXUALITY}+\text{TOUGH}+\text{EMPATHY}+\text{OPENNESS}+\text{ANXIETY} \\ &+ (1 + \text{ORDER} + \text{PLACE}|\text{SUBJECT})\end{aligned} \tag{3.2}$$

This model asks if listeners are more likely to look at the correct image during the sibilant interval, given the PLACE of articulation, CLUSTERTYPE, RETRACTIONCONDITION, and 20 ms TIMEWINDOW examined. Again, the social and cognitive predictors are included to account for potential interspeaker variability, but are not the primary area of investigation. Furthermore, the inclusion of by-subject random intercepts and by-subject random slopes for trial ORDER and PLACE of articulation suggest significant individual variability with respect to these predictors. Model convergence was not possible with by-subject random slopes or intercepts for CLUSTERTYPE, so they were pruned from the final model. The significant effects and interactions of the fixation accuracy model for preconsonantal sibilants are presented in Table 3.9.

Again, for the preconsonantal model, I will present the proportion of /sCr/ fixations although the model is fit on the accuracy of individual fixations. I will do so primarily because this allows for the visual representation of the different onset clusters to diverge precisely when the consideration of the candidates themselves diverge. Unlike in the prevocalic environments, remember that during the 180 ms window considered, the stimuli is still ambiguous. However, by the point at which rhotic or vowel onset is heard, the stimuli is disambiguated, whether that be to /sC/ or /sCr/. In Figure 3.6, a trial that is ultimately /sCr/

Table 3.9: Model predictions for all significant mains effects and interactions in fixation accuracy for /sC/ vs. /sCr/ clusters, N=26415. Place1 indicates the first contrast for Place, i.e. alveolar vs the combined velar and bilabial, and Place2 indicates the second contrast, i.e. velar vs. bilabial. A positive value indicates a greater prediction of fixations on the target word. Complete model predictions including variables and interactions that did not reach a significance threshold of 0.05 are included in the Appendix as Table A.2.

	Est.	*SE*	*t*	*p*
TimeWindow	0.11	0.03	2.89	0.003
ClusterTypeSCR	−0.23	0.05	−4.47	< 0.001
Hyper	−0.92	0.36	−2.52	0.016
TimeWindow:Increased	0.21	0.05	3.72	< 0.001
TimeWindow:Hyper	0.09	0.04	1.99	0.041
SCR:Place2	−0.31	0.13	−2.40	0.017
SCR:Increased	0.34	0.07	4.40	< 0.001
SCR:Hyper	0.43	0.07	6.33	< 0.001
Hyper:Place1	0.20	0.10	1.97	0.047
TimeWindow:SCR:Place1	0.19	0.06	3.12	0.002
SCR:Hyper:Place1	0.49	0.17	3.04	0.002

is indicated in teal and a trial that is ultimately /sC/ is indicated in red. The proportion of looks at the /sCr/ image is indicated on the y-axis. Like in the prevocalic environment, for trials presenting /sC/ and /sCr/ clusters, listeners begin with about one quarter of fixations to /sCr/ clusters while the remaining fixations are to /sC/ clusters, the fixation cross, or elsewhere on the screen. Thus in Figure 3.6, a correct fixation for /sCr/ clusters is indicated directly by a higher predicted /sCr/ fixation, while a correct fixation for /sC/ clusters may be indicated indirectly by a decreased predicted /sCr/ fixation.

In Figure 3.6, the fixation proportion is provided from the sibilant onset to one second after that onset. Again, vertical lines are provided as guideposts for what sound was heard when the eye movement was planned. If a look to an /sCr/ cluster was planned during the sibilant, it would be indicated in Figure 3.6 by an increased proportion of /sCr/ fixations between the dashed lines, which represent a 200 ms delay from the sibilant onset and stop onset respectively. A look to an /sCr/ cluster during the stop consonant, which are not manipulated in the present design and may contain disambiguating information, following

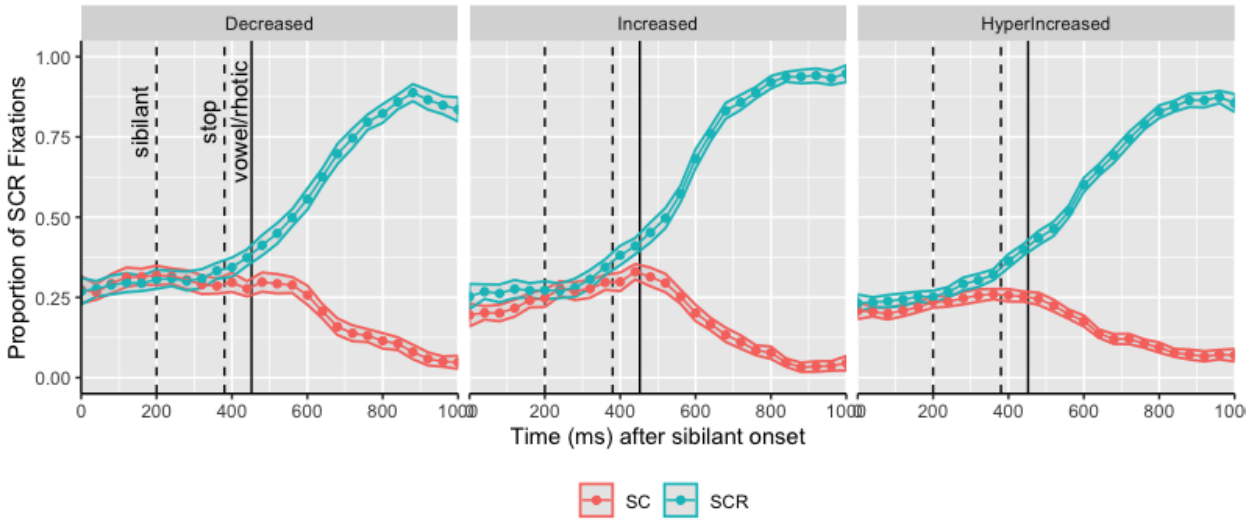


Figure 3.6: Fixation proportion for /sCr/ clusters (y-axis) by time following the sibilant onset (x-axis, binned into 40 ms windows), cluster type (color: /sC/ = red circles, /sCr/ = teal triangles), and retraction condition (panel row). The vertical lines represent a 200 ms delay from the onset of the sibilant, stop, and the vowel/rhotic mean for all clusters. A fixation executed during the sibilant interval would be observed between the black dashed vertical lines.

research on potential affrication of /t/ in /str/ clusters (Smith et al., 2019, see Section 2.3 for more), would be indicated by an increased proportion of /sCr/ fixations between the second dashed line, labeled 'stop', and the solid line, labeled 'vowel/rhotic'. A look once the stimuli has been disambiguated and the listener has heard the beginning of the vowel or rhotic is indicated by an increased proportion of /sCr/ fixations after the solid line.

Like in the prevocalic environments, preliminary inspection of Figure 3.6 may first highlight that the proportion of correct fixations does not rise dramatically until well after the rhotic/vowel onset. This suggests that, in many trials, listeners wait until the disambiguating information is available to identify the target word. However, turning to the sibilant interval between the dashed lines, I ask, can listeners use the spectral cues available during the sibilant interval in order to anticipate the upcoming rhotic? To test this we again look to steep slopes and diverging predictions for /sC/ and /sCr/ between the dashed lines. Visual inspection of the period between the sibilant and stop onset in Figure 3.6, indicates

that, across the different retraction conditions, the fixation proportion for /sCr/ clusters is increasing over time, as suggested by the positive slope for the teal lines, while no dynamic slopes are observed for /sC/ clusters in red. This is supported in part by the model with a main effect of TimeWindow ($t = 2.89, p = 0.003$), which predicts more correct fixations over the course of the sibilant interval. However, the two-way interaction of TimeWindow with ClusterType did not emerge as significant ($t = -0.65, p = 0.52$), suggesting that the proportion of correct fixations increases over time in both /sCr/ and /sC/ clusters, which simply may not be evident for /sC/ clusters as the proportion of /sCr/ fixations, rather than the proportion of correct fixations, is visualized in Figure 3.6.

Comparing the different RetractionCondition panels in Figure 3.6, visual inspection illustrates more positive slopes in the increased and hyper-increased retraction conditions as well as a divergence of the confidence intervals for /sCr/ and /sC/ clusters in the hyper-increased retraction condition. This is supported by the interaction of TimeWindow with RetractionCondition, with more correct fixations predicted over the course of the sibilant in the increased and hyper-increased retraction conditions compared to the decreased retraction condition (increased: $t = 3.72, p < 0.001$; hyper-increased: $t = 1.99, p = 0.041$). Furthermore, more correct fixations are predicted in /sCr/ clusters in the increased and hyper-increased retraction conditions compared to /sC/ clusters, as suggested by the interaction of ClusterType and RetractionCondition (increased: $t = 4.40, p < 0.001$; hyper: $t = 6.63, p < 0.001$). Taken together, these findings suggest that the spectral cues of retraction can improve correct fixations in /sCr/ clusters, well before the onset of the disambiguating rhotic.

Continuing an examination of fixation proportion in /sC/ and /sCr/ clusters, Figure 3.7 presents Place of articulation, ClusterType and TimeWindow. With Place of articulation, I ask if the spectral cues of retraction are better indicators of an upcoming /r/ in alveolar clusters, where /s/-retraction is more common and more advanced compared to

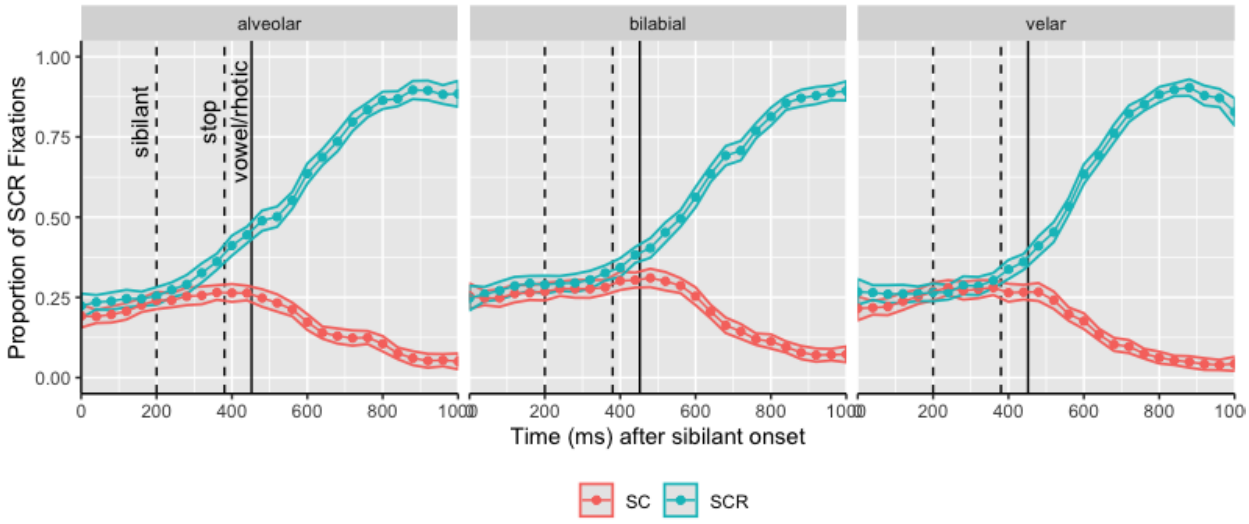


Figure 3.7: Fixation proportion for /sCr/ clusters (y-axis) by time following the sibilant onset (x-axis, binned into 40 ms windows), cluster type (color: /sC/ = red circles, /sCr/ = teal triangles) and place of articulation (panel column). The vertical lines represent a 200 ms delay from the onset of the sibilant, stop, and vowel or rhotic depending on the cluster identity. A fixation executed during the sibilant interval would be observed between the black dashed vertical lines.

velar and bilabial clusters. Visual inspection of the figure suggests that the proportion of correct fixations is greater in alveolar clusters than bilabial and velar clusters, as indicated by the higher proportion of /sCr/ fixations, the more positive slope for /sCr/ clusters, and the divergence of /sC/ and the /sCr/ confidence intervals, over the course of the sibilant in the left-most panel. This is not supported by the model as a main effect of alveolar PLACE (Place1: $t = -0.38, p = 0.703$) as an interaction of alveolar PLACE of articulation with TIMEWINDOW (Place1: $t = -0.69, p = 0.487$). PLACE of articulation did emerge as significant in its interaction with CLUSTERTYPE, but only for velar compared to bilabial clusters ($t = -2.40, p = 0.017$), not for alveolar (Place1: $t = 0.92, p = 0.357$). However, as we are concerned with the time course of correct fixations over the course of the sibilant interval, we can look to the significant interaction of TIMEWINDOW with CLUSTERTYPE and PLACE. This three-way interaction predicts more correct fixations over the course of the sibilant, but only for /str/ clusters (Place1: $t = 3.12, p = 0.002$; Place2: $t = -0.10, p = 0.917$). This

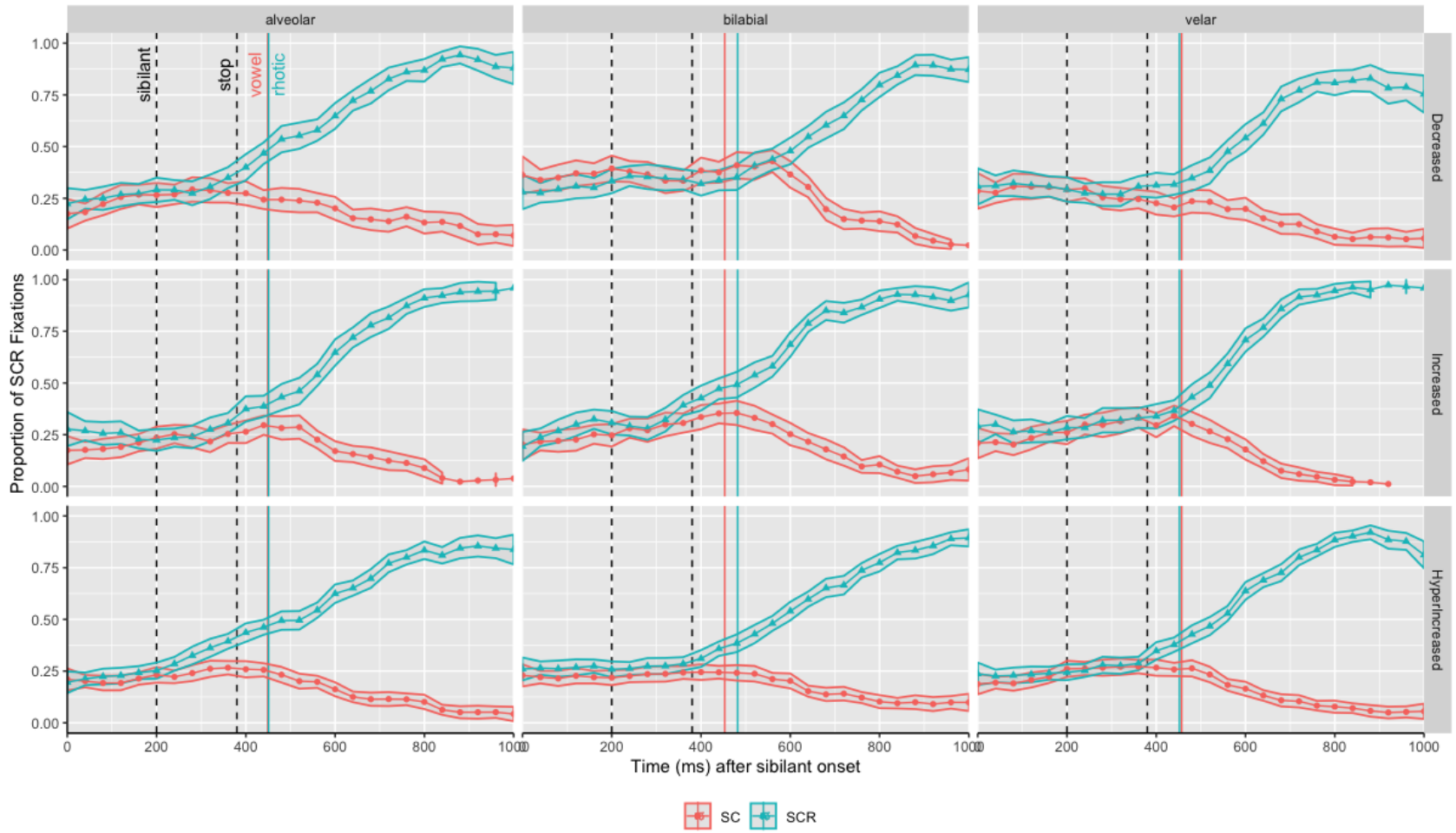


Figure 3.8: Fixation proportion for /sCr/ clusters (y-axis) by time following the sibilant onset (x-axis, binned into 40 ms windows), cluster type (color: /sC/ = red circles, /sCr/ = teal triangles), place of articulation (panel column), and retraction condition (panel row). The vertical lines represent a 200 ms delay from the onset of the sibilant, stop, and vowel or rhotic depending on the cluster identity. A fixation executed during the sibilant interval would be observed between the black dashed vertical lines.

suggests that listeners improve in their consideration of the correct candidate over the course of the sibilant only in /str/ clusters, precisely where retraction is the most available and the most expected.

Figure 3.8 puts all the pieces together, illustrating the proportion of /sCr/ fixations by both PLACE of articulation and RETRACTIONCONDITION. Visual inspection of the figure suggests, that while fixations on /sCr/ images are visibly increasing over the course of the sibilant for most pairings of PLACE and RETRACTIONCONDITION, only for alveolar clusters in the hyper-increased retraction condition do the confidence intervals for /sCr/ and /sC/ clusters not overlap. This is supported by two interactions in the model: firstly the interaction of PLACE and hyper-increased (compared to decreased) RETRACTIONCONDITION ($t = 1.97, p = 0.047$), which predicts more correct fixations for alveolar clusters in the hyper-increased retraction condition; and secondly, the interaction CLUSTERTYPE with alveolar PLACE of articulation and hyper-increased RETRACTIONCONDITION (Place1: $t = 3.04, p = 0.002$), which predicts more correct fixations for /str/ clusters specifically in the hyper-increased retraction. Taken together, along with the interaction of TIMEWINDOW, CLUSTERTYPE, and PLACE discussed just before, these findings suggest that in the environments where retraction is most expected and the greatest degree of retraction is observed, listeners look to the correct image more quickly than in environments were less retraction is expected or observed. Specifically, listeners are most able to immediately use the cues of /s/-retraction in /str/ clusters when the degree of retraction approaches a canonical /ʃ/.

The models and figures examined thus far present a combined fifty-two participants, which can potentially obscure individual differences in processing strategies. Before moving onto a discussion of the findings, I turn to an examination of the individual results in order to better understand how different individuals use the cues of /s/-retraction, especially given the between-subject design of this experiment. In particular, with this examination, I ask if some listeners in the decreased and increased conditions are able to use the cues of

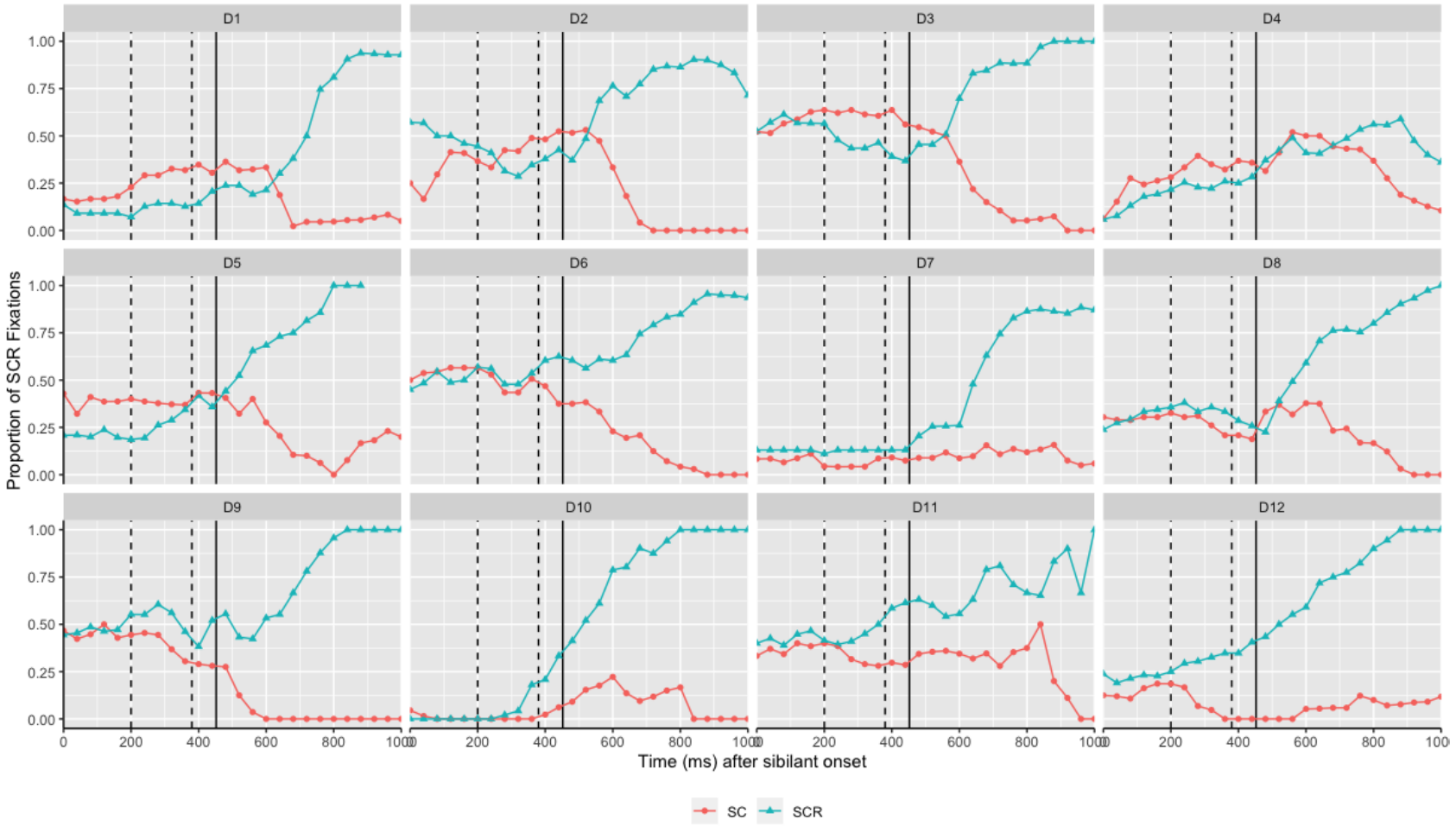


Figure 3.9: Fixation proportion for /sCr/ clusters (y-axis) by time following the sibilant onset (x-axis, binned into 40 ms windows), and cluster type (color: /sC/ = red circles, /sCr/ = teal triangles) in the decreased retraction condition for 12 representative participants. The vertical lines represent a 200 ms delay from the onset of the sibilant, stop, and vowel or rhotic depending on the cluster identity. A fixation executed during the sibilant interval would be observed between the black dashed vertical lines.

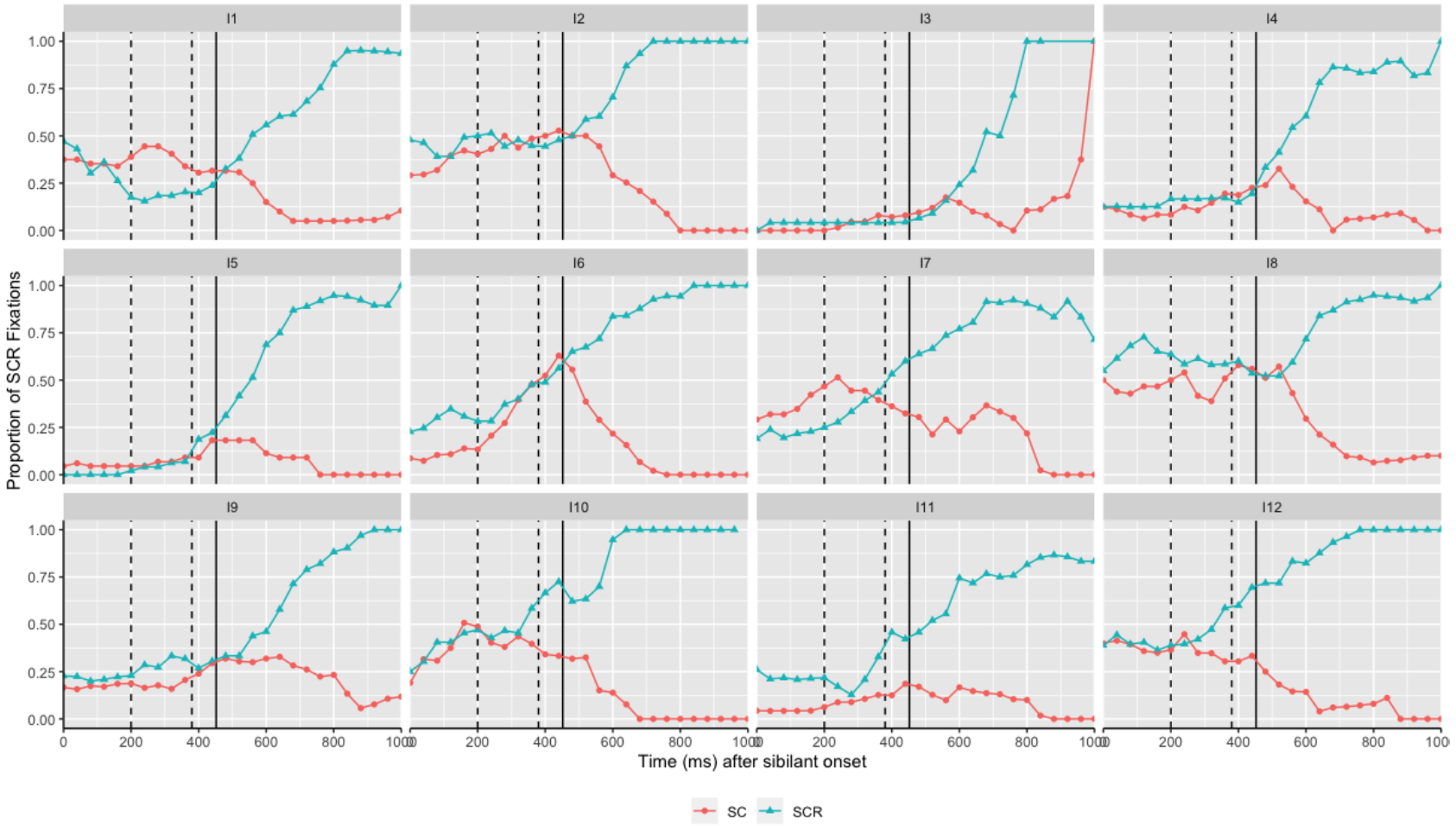


Figure 3.10: Fixation proportion for /sCr/ clusters (y-axis) by time following the sibilant onset (x-axis, binned into 40 ms windows), and cluster type (color: /sC/ = red circles, /sCr/ = teal triangles) in the increased retraction condition for 12 representative participants. The vertical lines represent a 200 ms delay from the onset of the sibilant, stop, and vowel or rhotic depending on the cluster identity. A fixation executed during the sibilant interval would be observed between the black dashed vertical lines.

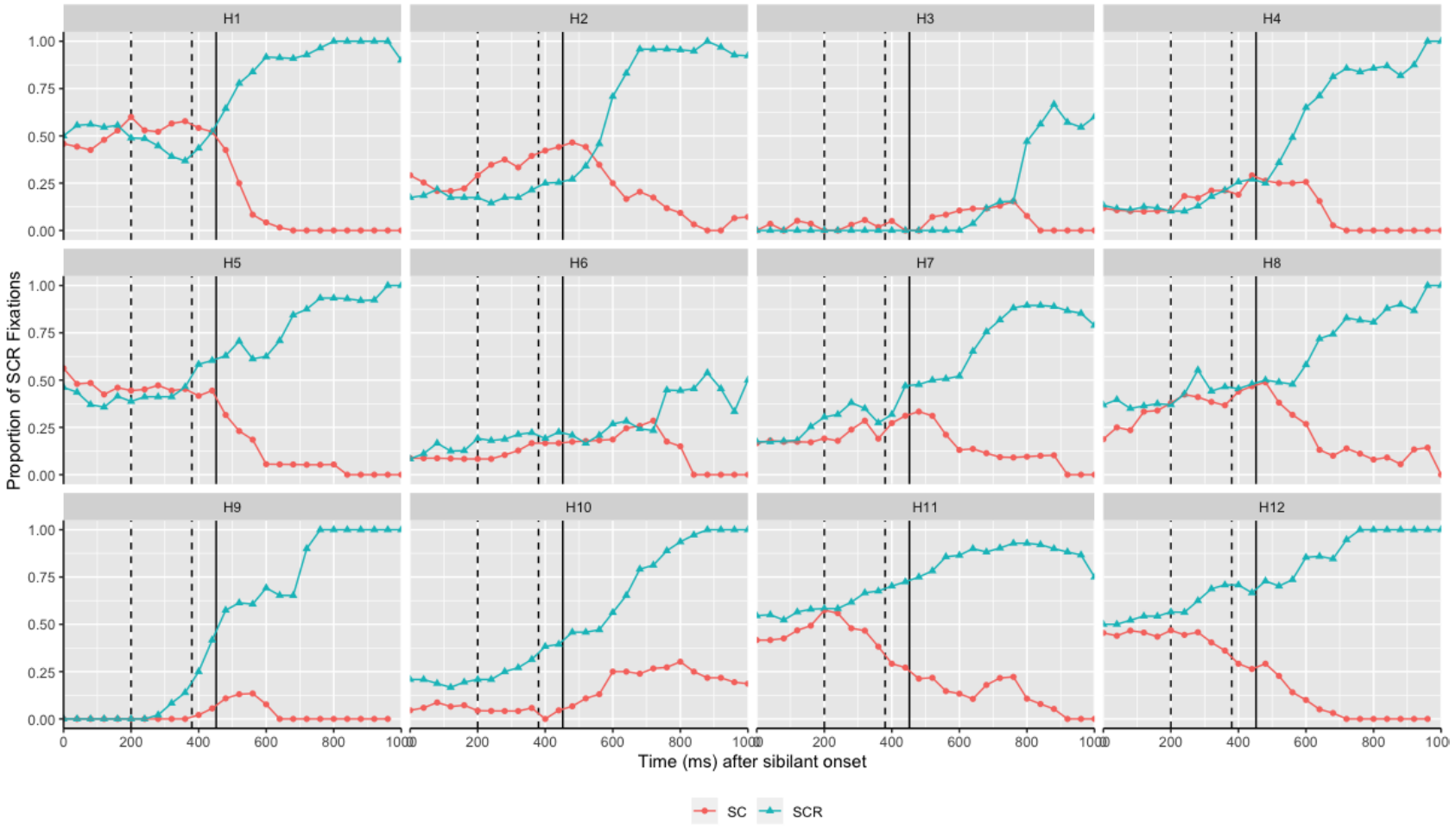


Figure 3.11: Fixation proportion for /sCr/ clusters (y-axis) by time following the sibilant onset (x-axis, binned into 40 ms windows), and cluster type (color: /sC/ = red circles, /sCr/ = teal triangles) in the hyper-increased retraction condition for 12 representative participants. The vertical lines represent a 200 ms delay from the onset of the sibilant, stop, and vowel or rhotic depending on the cluster identity. A fixation executed during the sibilant interval would be observed between the black dashed vertical lines.

retraction, which while less robust than the hyper-increased RetractionCondition, still provide varying degrees of contextual information. In Figures 3.9, 3.10, and 3.11, I present twelve representative participants from each of the decreased, increased, and hyper-increased retraction conditions respectively. Participants are ordered by the difference between their mean fixation proportions for /sCr/ and /sC/ clusters at 380 ms (or 200 ms after the stop onset). Although place was shown earlier to significantly influence an individual's likelihood of immediately integrating the spectral cues of retraction, the individual patterns presented here are combined across places to increase the number of observations and thus yield more reliable proportions. Visual inspection of the three figures highlights the diversity in speech processing patterns between listeners. For example, in each of three conditions there are some participants who have a fixation proportion of 0 on /sCr/ clusters at the sibilant onset, suggesting that are still looking at the fixation cross 873 ms later. Similarly, there are some participants who exhibit a fixation proportion of 50, suggesting that they always look to one image or the other before they have any potential disambiguating information. Yet the majority of participants exhibit a pattern similar to group results, where they have an approximately 25% chance of looking to the target, a 25% chance of looking to the competitor, and 50% change of looking to the fixation cross or the blank space on the screen.

In much the same vein, individual differences in fixation proportions are observed across the sibilant interval. In each RetractionCondition, there are participants more likely to exhibit correct fixations and participants less likely to exhibit correct fixations. However, in each RetractionCondition, the majority of participants appear equally likely to look to an /sCr/ image if the target word is /sCr/ or /sC/, despite the presence of a displayed confidence interval. While much of this variation does not appear to be systematic, one consistent pattern is that individuals who are less likely to look at either image before the onset of sibilant, i.e. a fixation proportion of 0 before the first dashed line, are unlikely to exhibit any noticeable differences in fixation proportion between the clusters by the end of

the sibilant, with one possible exception in the hyper-increased retraction condition (H9). Whether this is because these participants are simply slower at directing their eye gaze or because they are conscientious participants who believe that they should maintain gaze on the fixation point until they are certain of their response (thus putting themselves at a potential disadvantage) is unclear. Regardless, a high degree of individual variability is observed across all retraction conditions. This suggests that there are individuals who consistently use the coarticulatory cues of retraction, even when they're less available.

3.5 Discussion

The present experiment provides novel evidence for the perception of /s/-retraction. First and foremost, this experiment simply asks if listeners have context-dependent strategies of perception for preconsonantal sibilants. Specifically, using eye tracking during a lexical identification task, this experiment asks how quickly listeners can identify the intended target word depending on the relative expectation for /s/-retraction and degree of /s/-retraction presented. This experiment revisits the immediate integration of spectral cues for sibilants, following the findings of Galle et al. (2019) that listeners wait until the onset of the vowel before integrating the spectral cues of frication to distinguish prevocalic /s/ and /ʃ/. Additionally, this experiment builds of work by Beddor et al. (2013) finding that coarticulatory cues of vowel nasalization can be immediately integrated, asking if this is true for coarticulatory cues for longer distance dependencies like /s/-retraction.

Before examining the preconsonantal patterns, this experiment first asked if immediate integration of spectral cues is possible for prevocalic sibilants using the present design. Recall that Galle et al. (2019) found that listeners wait until the onset of the vowel before executing looks toward the target image when distinguishing prevocalic /s/ and /ʃ/. Galle et al. explore a variety of different explanations for this apparent buffer strategy, from the possibility that sibilants may not be interpreted as speech sounds until the vowel onset to

the possibility that spectral cues in frication are not reliable enough to facilitate immediate activation of one phonological category over another. The present experiment revisits cue integration for prevocalic sibilants in light of findings for preconsonantal sibilants, asking if immediate integration for sibilants is limited to preconsonantal environments, where no phonological contrasts exists and formant transitions are not available. The findings of the present experiment suggest that it is not, contrasting with Galle et al. to find evidence for immediate integration of spectral cues when considering prevocalic /s/ and /ʃ/ before the formant transition information is available. That is, as the prevocalic sibilant unfolds, listeners are more likely to look toward the correct image. In fact, comparing both the model predictions and data visualization, the evidence for cue integration is relatively stronger in prevocalic compared to preconsonantal environments, with more dramatic shifts that occur earlier.

It is not immediately clear how to reconcile the buffer strategy of speech perception observed for prevocalic sibilants by Galle et al. (2019) and the cascade strategy observed here for both prevocalic and preconsonantal sibilants. One potential explanation lies in differences in instructions between the studies: Participants in the present experiment were instructed to select the correct image as "quickly and accurately as possible", while Galle et al. "encouraged [participants] to take their time and perform accurately" (p. 12). It's possible that encouraging speed may encourage participants to immediately integrate cues that would otherwise be stored in buffer until additional cues become available.

Additionally, the different findings may stem from the different methods employed. The present study uses naturally-produced stimuli and presents two images: a target and a competitor. In contrast, Galle et al. use artificially generated sibilants and present four images: a target, a competitor, and two distractors. Galle et al. consider that their observed buffer strategy may be a consequence of their artificial stimuli, by which the listener may not have sufficient information to reliably make a contrast between /s/ and /ʃ/. They test this

hypothesis using a gating task, finding that listeners presented with only a portion of the frication can consistently and reliably categorize the onset as /s/ or /ʃ/. However, it is possible that the spectral information present in the artificially generated stimuli may be sufficient to make an ultimate categorization, but insufficient to be integrated in real time and influence consideration of the candidates. Galle et al. further test their observed buffer strategy by using naturally-produced sibilants, finding that listeners categorize stop consonants earlier than naturally-produced sibilants. The inclusion of the distractor candidates may also inhibit timely consideration of the correct candidate, as listeners don't have sufficient contextual information to use their context-dependent knowledge of /s/-retraction in real-time when additional candidates are viable. If this reasoning is correct, it is unclear what predictions this makes from real world conversations in which rich syntactic, semantic, and discourse context is available to the listener, but the no lexical items are explicit candidates. Another possibility is that the presence of the preconsonantal sibilants may have encouraged listeners to use a cascade strategy even in prevocalic environments. This may, following the unexpected effect of retraction condition aiding the processing of prevocalic sibilants, suggest that the presence of systemic and predictable context-dependent variation encourages listeners to use spectral cues as soon as they are available for all phonological contexts.

Finally, and perhaps most likely, the different findings may stem from the different analyses conducted by the two studies. The present study examines the proportion of correct fixations planned during the duration of the sibilant. In contrast, Galle et al. examine the relative bias in proportion of /s/ fixations after the bias departs from 0, which for in their study is approximately 150 ms after the vowel onset. Unlike the present study, Galle et al. are not asking whether consideration of the correct candidate can improve as a result of exposure to relevant cues, but rather at what point does the effect of the onset sibilant cross a threshold in biasing /s/ consideration. So while Galle et al. find that listeners are relatively

slower in categorizing a sibilant compared to a stop consonant, the present experiment finds that consideration of the correct candidate significantly improves during the sibilant itself.

Given that the analysis of the present experiment is limited to the 180 ms[3] temporal window offset 200 ms from the sibilant, in order to capture any looks planned during the the sibilant, it is worth noting that the size of the shifts observed is relatively small. The proportion of correct fixations does not reach 50%[4] until the onset of the vowel or the rhotic, after accounting for that 200 ms lag, and does not peak until approximately 600 ms after the vowel onset or 300-400 ms after the vowel/rhotic onset. Thus, it is not the case, either for the prevocalic or the preconsonantal sibilants, that listeners are making sudden, definitive categorizations during the sibilant, but rather that listeners are more likely to consider the correct candidate, as evidenced by their looking to the image representing that candidate, as the sibilant unfolds. This suggests that a cascade strategy of speech perception can be used for sibilants, but is not universally observed, as many participants wait until vowel or rhotic, holding the spectral cues in a buffer until additional cues are available.

Thus, with it established that listeners can use the spectral cues of sibilants immediately, this experiment asks if they use context-dependent strategies of processing those cues. Specifically, with Hypothesis 2, I ask if listeners can anticipate the presence of an upcoming /r/ from the spectral cues of frication alone, manifested acoustically by a lowered centroid frequency on the sibilant. In such a scenario, a listener who hears a more retracted onset sibilant may consider *string* to be a more viable candidate than *sting* and direct their gaze toward the image associated with that word, even before the /r/ has been perceived. The findings of the present experiment demonstrate that listeners are more likely to look to the correct image over the course of the sibilant, even when presented only with weak cues of

3. This in and of itself represents another distinction from Galle et al. (2019), whose sibilant stimuli were nearly double the length of those in the present experiment, at 350 ms. However, Galle et al. found no evidence to suggest that sibilant duration influences likelihood of immediate integrating spectral cues, nor has sibilant length been shown to contrast /s/ and /ʃ/ (McMurray & Jongman, 2011).

4. The chance that a given fixation falls within the area of the correct image is 23.08%.

retraction, illustrating that as spectral information unfolds, their consideration of the correct candidate improves. The observation that listeners can begin to distinguish /sC/ and /sCr/ clusters in all retraction conditions does not necessarily contradict this hypothesis, as even in the decreased retraction condition, there are still spectral cues of retraction in /str/ clusters, with a retraction ratio of 0.3 for /str/, but a retraction ratio of 0.1 for /st/ clusters. Hypothesis 2 is further bolstered when the onset sibilant in /sCr/ clusters is manipulated to be more retracted, with increasing looks to the correct image over the course of the sibilant in both the increased and hyper-increased retraction conditions compared to the decreased retraction condition. Taken together, these findings suggest that listeners are able to use the cues of /s/-retraction when considering lexical candidates, and do so as soon as that information is available.

Hypothesis 3 builds directly off Hypothesis 2, asking whether a more retracted /s/ is a better indication of an upcoming /r/ than a less retracted /s/ is for the absence of /r/. Recall that Beddor et al. (2013) found that vowel nasalization facilitated a greater anticipation of an upcoming nasal stop than an oral vowel did of an upcoming oral stop. The present experiment finds that listeners are more likely to look to the correct lexical candidate for /sCr/ than /sC/ onset clusters when the onset sibilant is manipulated to contain more spectral cues of retraction, i.e. in the increased or hyper-increased retraction condition. Like Beddor et al., this effect is not categorical, as listeners presented with a non-retracted sibilant did exhibit a small but consistent likelihood of considering the /sC/ image, albeit not the extent as /sCr/ clusters. These findings demonstrate that listeners attend closely to contextual cues when considering different candidates, but not all cues are equally informative. Specifically, these findings for /s/-retraction, paired with Beddor et al.'s findings for vowel nasalization, suggest that the presence of coarticulation is more informative than its absence.

Hypothesis 4 constitutes another follow-up to Hypothesis 2 by asking if the the cues of /s/-retraction are a better indication of an upcoming /r/ in /str/ clusters, where /s/-retraction

is most expected, than in /spr/ and /skr/ clusters, where it is less expected. The present experiment asks whether listeners listeners have detailed phonological knowledge about the distribution of /s/-retraction and, if they do, whether that detailed phonological knowledge influences the consideration of lexical candidates even as the acoustic signal unfolds. The findings of this experiment confirm this hypothesis, as listeners are more likely to look to /str/ clusters over the course of the sibilant, and more likely to do so when presented with the most extreme degrees of retraction. This was perhaps most clearly visible in Figure 3.8, where /str/ clusters in the hyper-increased retraction condition appear to show the greatest differentiation from /st/ clusters in the latter half of the sibilant. However, it is worth noting that there is a possible confound here: All /sCr/ clusters were not manipulated to contain the same degree of retraction in the same conditions, but rather were manipulated in reference to the model talker's natural production. In the hyper-increased retraction condition, /str/ clusters have a retraction ratio of 0.9, while /skr/ and /spr/ clusters have a retraction ratio of 0.7, both of may be perceived as /ʃ/ prevocalically. Thus, these findings cannot be interpreted to unequivocally state that the increased proportion of correct fixations in /str/ clusters is a result of the listener's expectations for retraction in those environments, and not a result of the simple higher degrees of retraction provided in those environments. Another possible confound is the effect of lexical frequency, as only in the alveolar place of articulation is the /sCr/ cluster more frequent than the /sC/ cluster. Thus, it is possible that all else being equal listeners are more likely look toward the image representing the more frequent word.

In summary, the findings of the present experiment demonstrate that listeners have detailed phonological knowledge about /s/-retraction and use that knowledge in real time in speech processing. Specifically, I have shown that listeners are more likely to look toward the correct candidate when there are more cues of retraction in /sCr/ clusters in general and in /str/ clusters in specific. These findings are particularly noteworthy in light of Galle

et al. (2019) who, after finding evidence for a buffer strategy in the perception of prevocalic sibilants, entertain the hypothesis that the highly variable and context-dependent nature of sibilants may impede the immediate integration of their spectral cues. Rather, the findings of the present experiment demonstrate that listeners are able to immediately integrate cues of that context-dependent variation in order to make predictions about the sibilant's phonological environment.

The finding that listeners are able to immediately integrate the cues of /s/-retraction when considering potential candidates has implications for /s/-retraction as a sound change. It demonstrates that the spectral cues on the /s/ serve as an important cue in contrasting /sCr/ and /sC/ sequences. This is not to say that the spectral cues on the sibilant are the primary cue contrasting these clusters; the results clearly demonstrate otherwise. That is, most looks to the correct image do not occur until after the clusters have been disambiguated by the clear presence or absence of the /r/. However, as /s/-retraction progresses, it is possible that listeners will begin to rely more heavily on the cues of /s/-retraction to contrast words like *string* and *sting*, particularly as more listeners reanalyze the onset to /ʃ/. In such a scenario, it is eventually possible that the spectral cues on the sibilant become the primary cue contrasting the clusters, making the subsequent /r/ redundant, and potentially lead to it being reduced or deleted entirely. In such a distant future, the contrast would not be between *string* and *sting*, but rather between *ʃting* and *sting*

CHAPTER 4

EXPERIMENT II: CATEGORIZATION

In the previous chapter (Experiment I: Cue Integration), I show that listeners have detailed phonological knowledge about preconsonantal sibilants and make use of that knowledge as soon as it is available. Like in Chapter 3, the present chapter examines the perception of preconsonantal sibilants, asking how individuals categorize ambiguous sibilants in these environments when forced to do so. Specifically, I ask, do they perceive a potentially ambiguous sibilant differently in /str/ clusters, where retraction is more expected due to both coarticulatory factors and its alignment with the ongoing sound change, than in /spr/ and /skr/ clusters, where the coarticulatory retraction is less observed? Furthermore, I ask whether different representations of masculinity influence the perception of these clusters in such a way that may shed light on the potential socio-indexical meaning of /s/-retraction in these environments. Finally, by recruiting a diverse listener pool, we can examine any potential shifts in apparent time between older and younger listeners that may illustrate how the perception of /s/-retraction has changed over time alongside the reported changes in production (e.g. Gylfadottir, 2015; Wilbanks, 2017), as discussed in the previous chapter (Section 2.5.1).

This experiment consists of a phoneme categorization task with nonce words in American English in which the onset sibilant in /sCr/ clusters is replaced with a step on a continuum from /s/ to /ʃ/, with more /ʃ/-like steps perceptually corresponding to an increased degree of retraction. Both the talker and the face are varied, manipulating the relative masculinity of the model talker. Listeners are asked to use a key press corresponding to orthographic representations of the nonce words presented on the screen, recording their categorization.

In this chapter, I first provide a brief background on the previous research in phoneme categorization, perceptual compensation, and attunement to social characteristics (Section 4.1). I then present an overview of the experiment (Section 4.2) and the materials and methods employed therein (Section 4.3). Next, I report the results (Section 4.4) and conclude with

a discussion of how listeners perceive ambiguous sibilants in preconsonantal environments.

4.1 An introduction to phoneme categorization

Listeners use a wealth of information available to them to aid in the perception and categorization of speech sounds. This includes not just acoustic information, but other linguistic information from phonological and prosodic to syntactic and semantic. Beyond that, listeners make use of a range of other information, including visual information about the articulators (e.g. the McGurk effect: McGurk & MacDonald, 1976), aerotactile feedback (Gick & Derrick, 2009), lexical frequency (e.g. the Gangong effect: Ganong, 1980), and social attributes of the speaker (Strand, 1999). Each of these components can influence perception such that an acoustically identical speech sound may be categorized differently in different contexts. In this section, I introduce the two sources of contextual information examined in this experiment: phonological, that is how listeners compensate for coarticulatory information (Section 4.1.1), and socio-indexical, that is how listeners make adjustments for perceived social attributes (Section 4.1.2).

4.1.1 Compensation for coarticulation

Perceptual compensation for coarticulation is the process by which listeners reduce or eliminate the acoustic effects of coarticulation in order to recover the intended target sound. A classic example from Mann & Repp (1980) examines the coarticulatory effects of vowel rounding on sibilant perception. Mann & Repp find that listeners are more likely to categorize an ambiguous sibilant as /s/ if it precedes /u/ than if it precedes /a/. This process first requires that listeners have experience with the coarticulatory influence that rounded vowels can exert on sibilants. This is realized articulatorily by the lip rounding from the vowel beginning during the sibilant, thus lengthening the cavity anterior to the constriction. These coarticulatory effects are realized acoustically by a general lowering in centroid frequency, leading

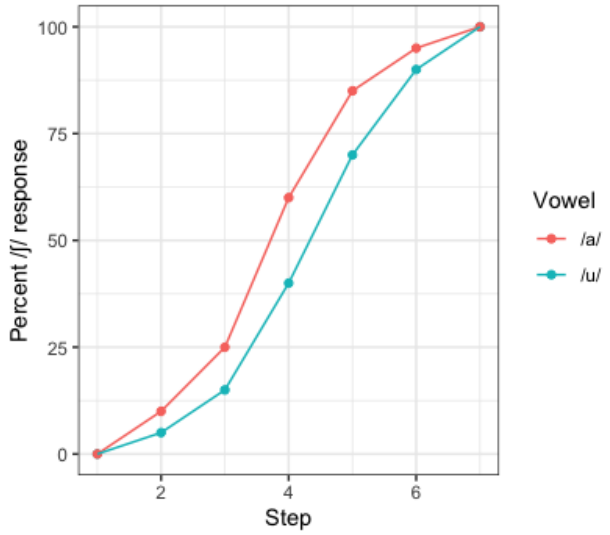


Figure 4.1: Example response curve using toy data to illustrate perceptual compensation in sibilant categorization (y-axis, as percent /ʃ/ response), by continuum step (x-axis, from /s/ to /ʃ/) and vocalic context (color: red = /a/, blue = /u/).

to an /s/ that sounds more /ʃ/-like. Secondly, listeners have to compensate for the coarticulation, meaning that they adjust their boundary for /s/ and /ʃ/ preceding rounded vowels to account for their experience with coarticulation in those environments. Consequently, an intermediary sound between /s/ and /ʃ/ is perceived as /ʃ/ preceding unrounded vowels and as /s/ preceding rounded vowels. Given that sibilants are generally perceived categorically, perceptual compensation is generally illustrated by a shifted response curve, seen in the toy example in Figure 4.1. In addition to sibilant-vowel coarticulation, perceptual compensation for coarticulation has also been observed for many other processes, including intrinsic pitch and vowel height (Hombert, 1977), vowel-to-vowel coarticulation (Beddor et al., 2002), and nasal coarticulation (Beddor & Krakow, 1999; Beddor, 2009), among others.

Perceptual compensation for coarticulation has been suggested by many researchers, chief among them Ohala (1993), as a possible path to sound change, proposing that sound change begins when listeners fail to compensate for extreme coarticulation. Instead, listeners encode a new speech target, which in turn influences later productions. Over time, the cues a listeners employed a phonological contrast shift from the coarticulatory trigger to the coar-

ticulatory target. For example, nasal stops exert coarticulatory influence over adjacent oral vowels, leading to variation from fully oral to highly nasalized. Over time, listeners may perceive the nasalized vowel as the intended target rather than the nasal stop, and eventually cease producing the nasal stop all together. Consequently, only the target (vowel nasalization) but not the trigger (nasal stop) remains. For /s/-retraction, the target is the sibilant and the trigger is the /r/. An Ohala model of sound change would predict that the contrast between historic /str/ and /st/ words would shift from the rhotic to the sibilant, and eventually be realized entirely by the sibilant, such that /ʃtɪŋ/ *string* contrasts with /stɪŋ/ *sting*.

While perceptual compensation for coarticulation has been robustly investigated, including much of the work presented in this section, less work has examined compensation for sound changes in progress that may test this hypothesis. In one notable exception, Harrington et al. (2008) examined /u/-fronting in Standard Southern British English. In this dialect, coronals and /j/ have traditionally exerted a great degree of coarticulatory force on /u/, leading to significant fronting relative to other environments. However, younger speakers appear to be fronting /u/ across the board, including in environments without the traditional coarticulatory triggers. In addition to examining participants' productions, Harrington et al. examine their categorization of /u/ in a traditionally fronting, i.e. coarticulatory, environment (*used*[1]*-yeast*) and a non-coarticulatory environment (*swoop-sweep*). From this categorization task, they observe that younger speakers compensate for the coarticulatory environment less than older speakers, and their phonological boundary for /i/-/u/ was shifted toward /i/ across the board in response to the sound change. These results, along with parallel findings for /u/-fronting in American English (Kataoka, 2009) and /ʊ/-fronting in British English (Kleber et al., 2011), empirically demonstrate that shifts in production and perception jointly advance in apparent time. However, it is unclear whether these find-

1. As in, "we *used* to go fishing."

ings directly speak to Ohala's hypothesis. In the /u/-fronting example, the sound change is extended to environments where the coarticulatory trigger is not present at all. However, in an Ohala hypothesis, the cue weight would shift from the trigger (the coronal) onto the target (/u/) and the trigger is subsequently deleted entirely. Following the pattern of sound change extension observed for /u/-fronting rather than a traditional Ohala model to sound change, /s/-retraction would be expected to extend to all instances of /s/, or perhaps all instances of preconsonantal /s/, even where no historic /r/ is present.

4.1.2 Adjustments for perceived speaker attributes

In addition to accounting for phonological context, listeners have been shown to attune to the social characteristics of the speaker. In much the same way as perceptual compensation for coarticulation, listeners may categorize an acoustically identical speech sound differently if they believe the speaker has a particular trait. In one classic example that again conveniently focuses on sibilants, Strand (1999) demonstrates that listeners account for the perceived gender of the speaker when categorizing a continuum from *sod* to *shod*. Listeners are aware that due to physiological and sociological reasons, men typically produce /s/ with a significantly lower centroid frequency than women. However, when presented with an acoustically identical, ambiguous sibilant, they categorize it as /ʃ/ if they perceive the speaker to be a woman and as /s/ if they perceive the speaker to be a man. This is because listeners make adjustments stemming from their knowledge and experience with speech in its social context. Furthermore, Strand demonstrates that listeners not only account for perceived speaker gender, but also for perceived gender typicality, being more likely to categorize the ambiguous sibilant as /s/ if they perceive the male talker to be more stereotypically masculine, or as /ʃ/ if they are perceived to be less masculine. This finding illustrates that adjustments are not only made for more categorical, broad social groups, but also for nuanced, performative traits and characteristics. Furthermore, listeners' attitudes about the relevant social charac-

teristics can strengthen the effect of these social factors. Building off the findings of Strand (1999), Campbell-Kibler (forthcoming) finds that in a *sod–shod* categorization task, listeners' adjustments for perceived gender and gender typicality is influenced by their relative endorsement of gender stereotypes, as determined from a short survey of questions on their opinions about traditional gender roles (MRAS: Pleck et al., 1993). Also using the MRAS survey, Levon (2014) found that listeners' attitudes about traditional gender roles influence the social meaning they assign to linguistic variables like sibilant production.

4.2 Study overview

Experiment II: Categorization asks if listeners perceive /s/-retraction differently in /str/ clusters, where it is expected, versus /spr/ and /skr/ clusters, where it is relatively unexpected. More specifically, this experiment asks if listeners make adjustments in the categorization of the onset sibilant in /str/ clusters in order to account for their experience with /s/-retraction in that phonological environment, i.e. do listeners compensate for /s/-retraction. However, as English has no phonotactic contrast between the sibilants /s/ and /ʃ/ preconsonantally, the categorization could not be done with lexical words, necessitating the use of nonce words. In this experiment, listeners heard nonce words with /SCr/ onsets, where /S/ represents any step on the continuum from /s/ to /ʃ/, and were asked to categorize the word as beginning with ⟨s⟩ or ⟨sh⟩.

The nature of the task allows for the inclusion of additional research questions on the perception of /s/-retraction. As discussed in Section 2.6, there are conflicting predictions for the potential social meaning of /s/-retraction. Recall that work on prevocalic sibilants has found that a retracted /s/ is generally evaluated as more masculine (e.g. Campbell-Kibler, 2011a) and listeners attune for performances of masculinity in their perception of prevocalic sibilants (e.g. Strand, 1999). However, research in /str/ clusters has not found a significant influence of retraction on social evaluation (Phillips, 2018). On the other hand,

online meta-commentary suggests that, for some individuals, retraction in /str/ clusters is an indicator of masculine toughness and straightness. In this categorization task, by varying the talker and co-present pictures of faces, we can ask how listeners additionally make adjustments for performances of masculinity, which speak to the potential socio-indexicality of /s/-retraction in these consonant clusters. We can also ask how listeners' opinions about masculine stereotypes and norms influence their categorization.

The simplicity of the task allows for it to be run using an online platform that can reach diverse listeners across the country. Without the limitations of university subject pools, we can examine how listener age influences their perception of /s/ in these environments, which may illustrate apparent time trends in the perception of /s/-retraction that parallel the observed apparent time trends for the production of /s/-retraction, as described in Chapter 2.5.1.

In Section 4.3, I outline the methods and materials used, including stimuli creation (4.3.1), participants and procedure (4.3.2), and analysis and hypotheses (4.3.3). In Section 4.4, I present the results of this experiment, and in Section 4.5, I move onto a discussion of the findings.

4.3 Methods & materials

4.3.1 Stimuli materials

Due to the phonotactic restrictions of English, it was not possible to do a phoneme categorization task with lexical words, like the *sod-shod* continuum used by Strand (1999) and Campbell-Kibler (forthcoming). For this reason nonce words were selected, allowing the potential categorization of both /s/ and /ʃ/ preconsonantally.

To select the nonce words and their orthographic representations, I first created eight paradigms of sibilant-initial nonce words in which each paradigm contained /s/, /spr/, /str/,

/skr/, /ʃ/, /ʃpr/, /ʃtr/, and /ʃkr/ onsets. For example, one such paradigm was /sutsi/, /sprutsi/, /strutsi/, /skrutsi/, /ʃutsi/, /ʃprutsi/, /ʃtrutsi/, and /ʃkrutsi/. Eight stop-initial paradigms were included as fillers that contained /gw/ clusters that are illicit in English but marginally acceptable to native speaker and found in loanwords, like the name *Gwen* /gwɛn/. In this way, the /ʃCr/ clusters were not the only marginally illicit clusters in the pilot. A 19-year-old female from Ohio recorded each of the paradigms at 48,000 Hz in an isolated double-walled sound booth with a Zoom H6 recorder and a Shure SM10A head-mounted microphone. One sibilant and one stop-initial paradigms were inconsistently pronounced by the research assistant and were thus eliminated as potential candidates for the categorization task.

In a small pilot, eight undergraduate students were provided with a pen and a response sheet and asked to give a potential English spelling for each of the nonce words played binaurally. From this pilot, it emerged that the participants prefer ⟨sh⟩ rather than the ⟨sch⟩ spelling common in some Yiddish/German loanwords like *schtick* and *schnapps*. Additionally, the participants preferred ⟨c⟩ over ⟨k⟩ in /ʃkr/ clusters. One paradigm (*sprimble, shprimble, strimble, shtrimble, scrimble, shcrimble*) was most consistently spelled by all eight participants and was thus selected as the stimuli for Experiment II.

Two white, male speakers from Iowa (age 19 and 21) were recruited to serve as the model talkers for this experiment and received payment or credit for an introductory linguistics course for their participation. All target and filler words were presented in the carrier phrase: *Please say X again: X X X* and recorded in an isolated double-walled sound booth with a Zoom H6 recorder and a Shure SM10A head-mounted microphone. The model talkers were asked to read the phonotactically licit nonce words with /sCr/ onsets (*sprimble, strimble, scrimble*), as well as the equivalent nonce words with the simplex /s/ and /ʃ/ onsets (*simble, shimble*) and various stop-initial nonce words (e.g. *kittle, quittle, gittle, gwittle*). They were not asked to read the marginally illicit /ʃCr/ onsets (*shprimble, shtrimble, shcrimble*).

Figure 4.2: Sample faces for categorization task from the Chicago Face Database (Ma et al., 2015). The image on the left (WMT1) normed as more masculine than the face on the right (WMN3).

To create a continuum from /s/ to /ʃ/ for the phoneme categorization task, the sibilant onsets from the prevocalic equivalents, *simble* and *shimble*, were extracted and digitally mixed using a modified Praat script originally created by (Darwin, 2005) at seven scaling ratios: 95%-/s/:5%-/ʃ/, 80%-/s/:20%-/ʃ/, 65%-/s/:35%-/ʃ/, 50%-/s/:50%-/ʃ/, 35%-/s/:65%-/ʃ/, 20%-/s/:80%-/ʃ/, and 5%-/s/:95%-/ʃ/. Each of the seven steps was cross-spliced onto the preconsonantal target word, creating a continuum from /s{p,t,k}rɪmbəl/ to /ʃ{p,t,k}rɪmbəl/.

The auditory stimuli from both model talkers were paired with eight images of faces from the Chicago Face Database (Ma et al., 2015) that were normed as more or less masculine than average. All faces used were white male faces that were consistently rated by participants in the norming study as being perceived as white, between ages 18 and 30, and of average level of attractiveness. However, by controlling for level of attractiveness, the relative difference in perceived masculinity between the more and less masculine talkers was rather small due to the general trend for more gender typical faces to be perceived as more attractive. Two sample faces are provided in Figure 4.2 and the perceived age, attractiveness rating, and masculinity rating for the target faces are provided in Table 4.1.

Table 4.1: Selected faces and their normed perceived ratings on age, attractiveness, and masculinity. A higher value indicates a face rated as more attractive/masculine.

More masculine faces			
code	age	attractiveness	masculinity
WMT1	25.12	3.05	4.66
WMT2	27.79	3.46	4.61
WMT3	22.17	3.96	4.59
WMT4	18.73	3.27	4.59
Less masculine faces			
code	age	attractiveness	masculinity
WMN1	20.39	3.54	3.61
WMN2	21.50	2.97	3.48
WMN3	25.12	3.12	3.67
WMN4	18.73	3.08	3.46

4.3.2 Participants & procedure

Participants

200 participants were recruited online through Amazon Mechanical Turk and were paid for their participation (at $12 per hour). All participants were using an internet connection located in the U.S. and reported being both born in the U.S. and raised in a household using English as its primary language until age 12, which were collectively used as a metric for determining native American English status. The mean age of the participants was 34.5, with a minimum age of 19 and a maximum of 60. The participant pool was skewed toward self-identifying male (120 male, 80 female, 0 non-binary), white (161 white, 14 black, 8 Asian, 8 Hispanic, 4 Native American, 7 other) and straight (185 straight, 7 gay or lesbian, 7 bisexual, 1 asexual). Participants exhibited significant diversity in geographic environment (59 urban, 93 suburban, 48 rural) and distribution (See Figure 4.3). An additional 38 individuals participated but were excluded from analysis due to non-native status, non-attentive responses, or self-reported speech/hearing disorders.

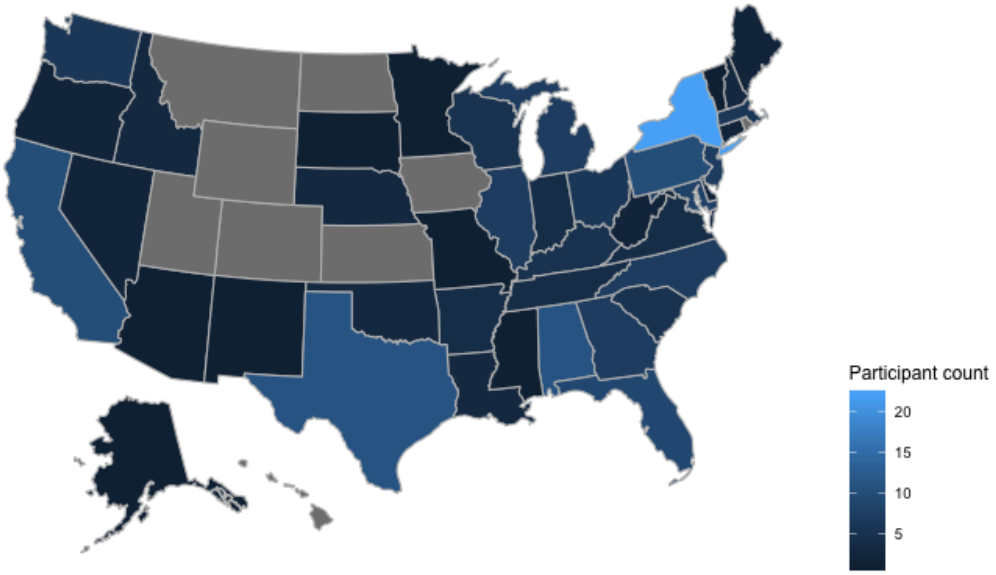


Figure 4.3: Geographic distribution of participants by self-reported state lived in the longest before age 18, with a lighter blue indicating a higher count of participants. Gray indicates no participants reporting living in that state.

Procedure

Participants heard 42 trials (7 steps × 3 consonant clusters × 2 model talkers = 42 trials), plus fillers interspersed to check for attentiveness. Participants were randomly assigned to a condition that included either faces normed as more masculine for both talkers, less masculine for both talkers, or one more and one less masculine face. Participants were instructed that the faces presented were pictures of the talkers. After each trial, participants responded with a key press corresponding to the orthographic representations presented on the screen, e.g. *sprimble* or *shprimble*. A sample trial in Figure 4.4 illustrates how the visual stimuli and response options were presented.

Following the phoneme categorization task, participants evaluated the faces and voices presented in the experiment on a nine-point Likert scale for masculinity, toughness, and attractiveness. Participants also completed a short survey containing basic demographic questions, as well as the Male Role Attitude Scale (MRAS, Pleck et al., 1993). This survey

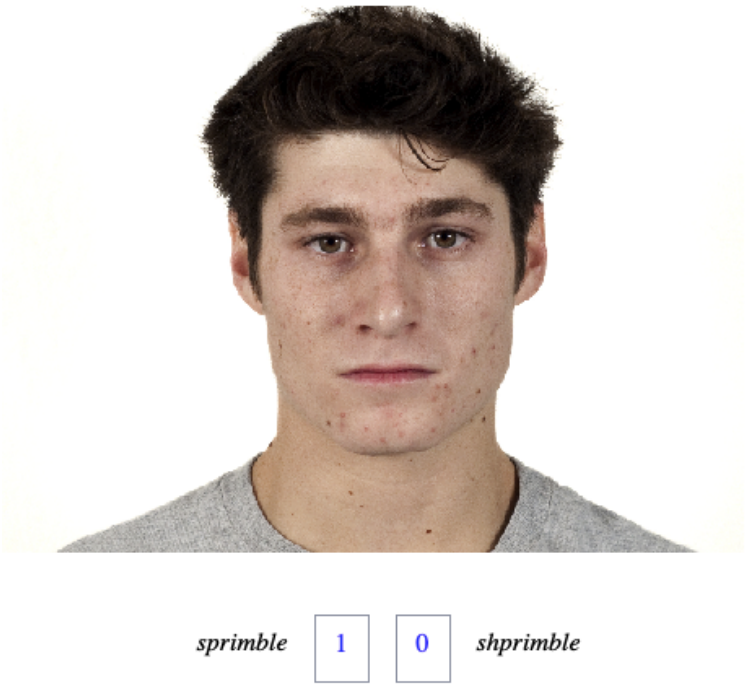


Figure 4.4: A sample trial illustrating the presentation of the prompt, photograph of the face, and response options.

is included to ask whether individuals' categorization strategies are influenced not just by community-defined measurements of masculinity, like the rating of the faces and voices, but also by the extent to which participants hold or endorse those community-defined measurements to begin with. Participants for the present experiment had a mean composite score of 27.95 (s.d. 6.26; on a scale of 10 to 40), where a higher score indicates a stronger endorsement of traditional stereotypes of masculinity. On the toughness subscale, participants had a mean score of 5.95 (s.d. 1.51; on a scale of 2 to 8), where a higher score indicates a stronger endorsement of toughness stereotypes. As a whole, participants for the present experiment were approximately normally distributed on the composite score with the mean just above the median possible value of 25, indicating that on average, participants are slightly more likely to endorse masculine stereotypes than to reject them. On the toughness subscale,

participant responses were skewed toward endorsing, rather than rejecting, masculine stereotypes of toughness. The participants for the present experiment were much less homogeneous than those in Experiment I: Cue Integration (Chapter 3), with a mean composite score of This is due to the relative heterogeneity of the participant population (Amazon Mechanical Turk workers online across the U.S.) compared to the previous experiment (students at the University of Chicago).

4.3.3 Hypotheses & analysis

Analysis

Listeners' responses (/s/=0, /ʃ/=1) were modeled with logistic mixed effects regressions using the `glmer()` function in the `lme4` package (Bates et al., 2015) in `R` (R Core Team, 2015). The fixed effects predictors included in the model were trial order (ORDER: 1–42, scaled), continuum step (STEP: 1–7; scaled), onset cluster identity (CLUSTER: /Str/, /Spr/, /Skr/; Helmert-coded[2]), participant's age (AGE: 19–60; log transformed to approximate normal distribution: 2.94–4.10; scaled), participant's toughness rating for the voice (TALKERTOUGHNESS: 1–9; scaled), the participant's toughness rating for the displayed face (FACETOUGHNESS: 1–9; scaled), and the participant's relative endorsement of masculine stereotypes of toughness from the MRAS survey (TOUGHNESSENDORSEMENT: 2–8; scaled). Other self-reported demographic information, including participant gender, sexuality, and location, did not improve model likelihood and were not included in the final model.

A preliminary model was designed with all two- and three-way interactions between the fixed effects predictors, and all interactions that did not reach a significance threshold of 0.05 were pruned from the final model. Additionally, a preliminary model included a maximally

2. Helmert-coding compares each level of a factor to the mean of the subsequent levels. Here, the first contrast compares /Str/ to the mean of the other clusters (/Spr/ and /Skr/) and the second contrast compares /Spr/ to /Skr/.

specified random effects structure but failed to converge. The model reported is a result of progressive simplification of the random effects structure until convergence was reached, which includes random intercepts for participant with by-subject random slopes for Order and Cluster. The fit for the logistic mixed effects model in `lme4` format is provided in 4.1.

$$\begin{aligned}&\textsc{Response} \sim \textsc{Order} + \textsc{Step*Age*ToughnessEndorsement}+\\&\textsc{Step*Cluster*ToughnessEndorsement} + \textsc{Step*Cluster*FaceToughness}+\\&\textsc{Step*Cluster*TalkerToughness} + (1 + \textsc{Trial} + \textsc{Cluster}|\textsc{Subject})\end{aligned} \tag{4.1}$$

Hypotheses

The specific hypotheses for listeners' categorization patterns to be tested are:

Hypothesis 1 The first hypothesis proposes that listeners attend to phonological environment when categorizing sibilants. Specifically, this hypothesis predicts that listeners will compensate for coarticulation in /Str/ clusters relative to /Spr/ and /Skr/ clusters, perceiving the onset sibilant as /s/ at more /ʃ/-like steps. If confirmed, this would demonstrate that listeners have detailed phonetic knowledge about /s/-retraction as a sound change, with greater degrees of retraction observed in /str/ clusters (Baker et al., 2011), and adjust their perceptual strategies to account for that knowledge. This would mirror findings for perceptual compensation for vowel quality in prevocalic sibilants (Mann & Repp, 1980) and coarticulatory vowel nasalization (Beddor & Krakow, 1999), among other phenomena. A confirmation of this hypothesis would be realized by fewer /ʃ/ categorizations at higher, i.e. more /ʃ/-like, steps for /Str/ clusters relative to /Spr/ and /Skr/ clusters.

Hypothesis 2 The second hypothesis predicts that there is an apparent time trend in categorizing preconsonantal sibilants. This follows from the observation of time trends in /str/ production (Gylfadottir, 2015; Wilbanks, 2017; Smith et al., 2019) and suggests that a parallel trend can be observed in perception. Specifically, this hypothesis predicts that younger speakers will compensate for coarticulation *more* than older speakers. This contrasts with the /u/-fronting example explored in Section 4.1.1 (Harrington et al., 2008). For /u/-fronting, older older speakers exhibit coarticulatory fronting of /u/ following coronals and compensate for coarticulation in those environments, while younger speakers front /u/ across the board and thus do not compensate for coarticulation in coronal environments. Here, however, younger speakers are predicted to compensate more rather than less, as /s-retraction has not extended to environments in which the coarticulatory trigger is not present. A confirmation of this hypothesis would be realized by fewer /ʃ/ categorizations at higher, i.e. more /ʃ/-like, steps for younger listeners relative to older listeners.

Hypothesis 3 The third and final hypothesis proposes that listeners attune to performances of masculinity in their categorization of preconsonantal sibilants. Specifically, this hypothesis predicts that listeners attune to the masculinity of the face and voice presented and attribute /s/-retraction to performances of masculinity when it aligns with the presented face, in the same way that they attune for masculinity in prevocalic sibilants (Strand, 1999). Furthermore, this proposes that listeners are influenced by their own relative endorsement of masculine stereotypes in their likelihood of attributing retraction to a performance of masculinity, as in prevocalic sibilants (Campbell-Kibler, forthcoming). A confirmation of this hypothesis would demonstrate that a retracted /s/ in /sCr/ environments indexes masculinity in much the same way as in prevocalic environments. A confirmation of this hypothesis would be manifested by fewer /ʃ/ categorizations at higher, i.e. more /ʃ/-like, steps for more masculine faces and voices and listeners who more strongly endorse masculine stereotypes.

Table 4.2: Model predictions for all main effects and interactions in sibilant categorization in different phonological environments and with different indicators of masculine stereotypes of toughness, N=18476. A positive value indicates stronger /ʃ/ prediction. Cluster1 indicates the first contrast for Cluster, i.e. /Str/ vs the combined /Spr/ and /Skr/, and Cluster2 indicates the second contrast, i.e. /Spr/ vs. /Skr/. Complete models predictions including variables and interactions that did not reach a significance threshold of 0.05 are included in the Appendix as Table A.3.

	Est.	*SE*	z	p
(Intercept)	−0.69	0.07	−10.72	< 0.001
Step	1.00	0.02	41.03	< 0.001
Cluster2	0.20	0.09	2.31	< 0.05
FaceToughness	0.09	0.04	2.28	< 0.05
Step:Age	0.19	0.02	7.82	< 0.001
Step:Cluster1	0.19	0.02	7.79	< 0.001
Step:Cluster2	0.11	0.01	7.94	< 0.001
Step:ToughnessEndorsement	−0.09	0.02	−4.07	< 0.001
Step:FaceToughness	−0.05	0.02	−2.24	< 0.05
Cluster1:TalkerToughness	0.09	0.02	2.67	< 0.001
Step:Cluster1:ToughnessEndorsement	−0.05	0.02	−1.99	< 0.05
Step:Age:ToughnessEndorsement	0.12	0.03	4.85	< 0.001

4.4 Results

The fit for the logistic mixed effects model is presented in Table 4.2. The intercept of the model emerged as significant ($z = -10.72, p < 0.001$), suggesting that all else being equal, listeners are more likely to categorize a preconsonantal sibilant as /s/ than /ʃ/. In fact, nearly two thirds of all responses were /s/, with 11,427 trials categorized as /s/ and 6,801 as /ʃ/. This result may be interpreted as a consequence of the English phonotactics, which only permit /s/ preconsonantally and thus neutralizes the phonological contrast between /s/ and /ʃ/ in these environments. This may allow participants to answer /s/ at significantly higher rates than in prevocalic sibilant categorization tasks. Relatedly, this effect may be attributed to English orthography, with the ⟨sCr⟩ sequences more common and acceptable than the ⟨shCr⟩ sequences.

There was no observed effect for trial Order, suggesting that listeners are no more or less

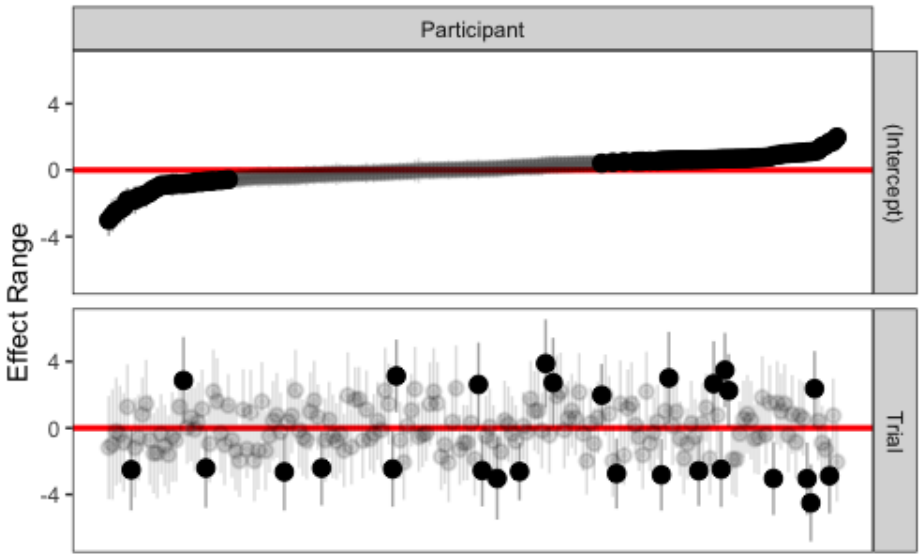


Figure 4.5: Participants' intercepts (top panel) and random effects for Trial Order (bottom panel). Points distinguishable from 0 (the red horizontal line) are highlighted (boldface).

likely to categorize the sibilants as /ʃ/ as the experiment progresses. However, the inclusion of by-subject random slopes for trial order suggests a high degree of individual variability in this effect. Individual intercepts and random slopes for trial Order are illustrated in Figure 4.5. As illustrated in the bottom panel, there was a wide variation in by-subject slopes for trial order, including participants who increased their likelihood of responding /ʃ/ over the course of the experiment (as seen by the bolded points above the red line) and participants who decreased their likelihood of responding /ʃ/ (as seen by the bolded points below the red line).

A main effect of Step was observed ($z = 41.03, p < 0.001$), which illustrates an increase in /ʃ/ responses as the scaling ratio of /ʃ/ increases. This is demonstrated in the left-hand panel of Figure 4.6, with a greater proportion of /ʃ/ responses at higher, i.e. more /ʃ/-like steps. The interaction of Step and Cluster is illustrated in the right-hand panel of Figure 4.6. Visual inspection of the figure indicates that more /ʃ/ responses are predicted at higher, i.e. more /ʃ/-like, steps for /Spr/ and /Skr/ relative to /Str/ clusters. No visually detectable

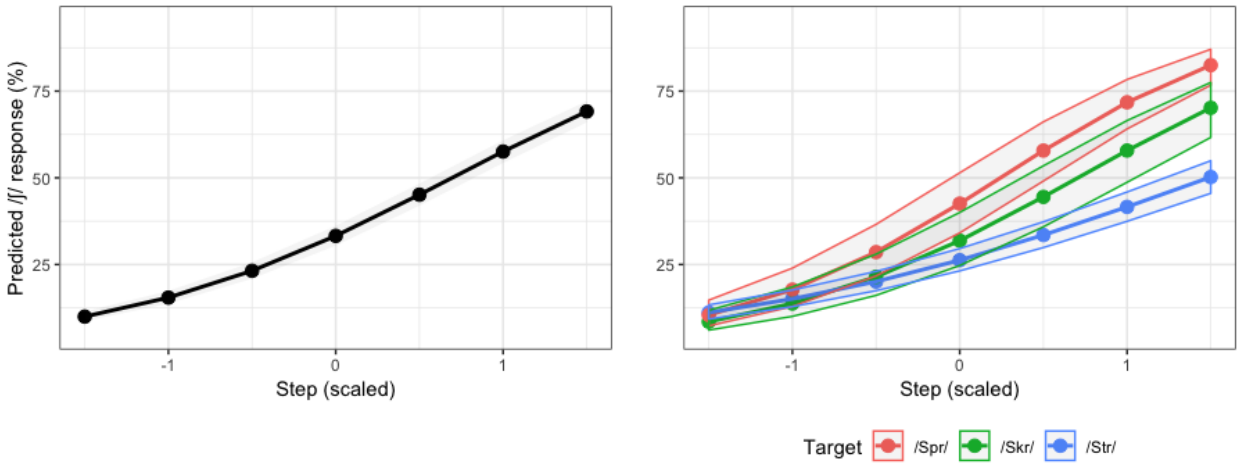


Figure 4.6: Left: Predicted /ʃ/ response (y-axis) by continuum step (x-axis). Right: Predicted /ʃ/ response (y-axis) for the interaction of continuum step (x-axis) and consonant cluster (color: red = /Spr/, green = /Skr/, blue = /Str/).

differences in responses between the clusters can be observed at lower, i.e. more /s/-like, steps. This interaction emerged as significant in model, with fewer /ʃ/ responses predicted at higher steps in /Str/ clusters compared to /Spr/ or /Skr/ clusters ($z = 7.79, p < 0.001$). Additionally, fewer /ʃ/ responses were predicted at higher steps in /Skr/ clusters compared to /Spr/ clusters ($z = 7.94, p < 0.001$). These findings suggest that listeners are the least categorical in their perception of /Str/ clusters and the most categorical in the perception of /Spr/ clusters.

This experiment examines listener age to ask if the apparent time findings for /s/-retraction production can also be observed in perception. No significant main effect of Age was observed, illustrating that older participants are not categorizing the sibilants differently across the board than younger listeners. This suggests that, all else being equal, there are no apparent time shifts in the phonotactic bias against /ʃ/ in these onset clusters. However, as the interaction of Step and Age illustrates in Figure 4.7, younger listeners appear to give fewer /ʃ/ responses than older listeners at higher, i.e. more /ʃ/-like, steps and slightly more /ʃ/ responses at lower, i.e. less-/ʃ/-like, steps. The interaction of Step and Age in the model supports this observation ($z = 8.14, p < 0.001$), suggesting that younger

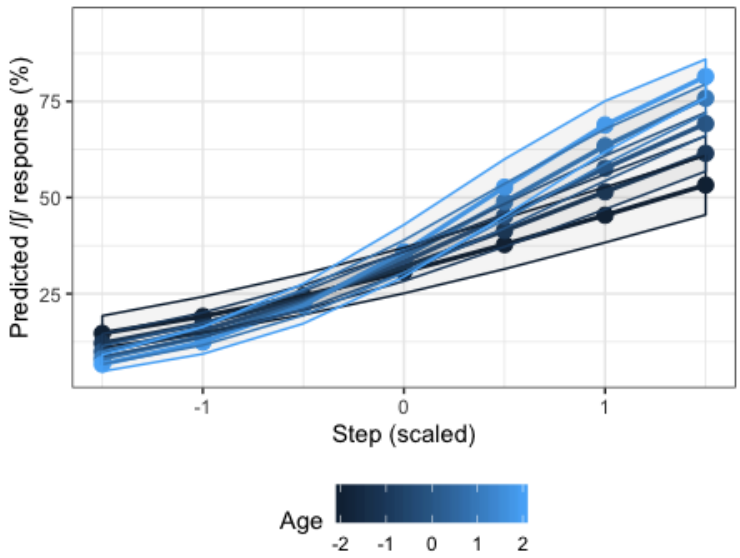


Figure 4.7: Predicted /ʃ/ response (y-axis) for the interaction of continuum step (x-axis) and age (color: darker blue = younger, lighter blue = older). Age is log-transformed and binned for ease of visualization but analyzed continuously.

listeners are less categorical in their perception of onset sibilant clusters than older listeners. The interaction between Cluster, and Age, as well as the three-way interaction between Step, Cluster and Age, did not reach the level of significance and were pruned from the final model, suggesting that younger listeners are not accounting for potential differences in coarticulation between the clusters more or less than older listeners.

This experiment asks what role masculine stereotypes play in the phoneme categorization of preconsonantal sibilants and how that may differ from prevocalic environments. Three different indicators of masculine stereotypes were examined: the model talker's voice (TalkerToughness), the face presented (FaceToughness), and the participant's relative endorsement of masculine stereotypes of toughness (ToughnessEndorsement). All faces normed by the Chicago Face Database (Ma et al., 2015) as more masculine were evaluated by our participants as tougher and more masculine than the faces normed as less masculine. One of the two model talkers was also consistently evaluated by our participants as tougher and more masculine than the other. Listeners' relative endorsement of masculine

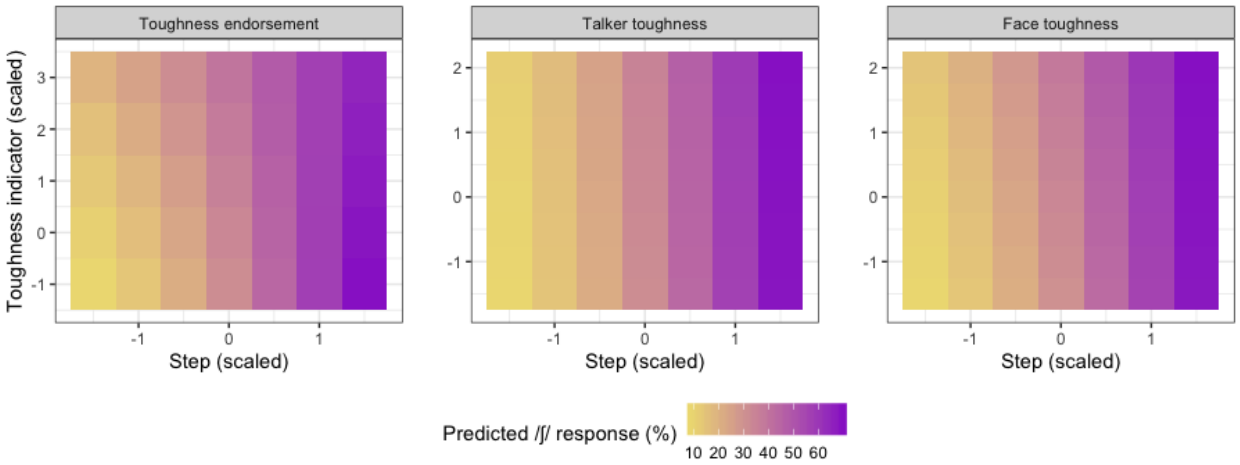


Figure 4.8: Predicted /ʃ/ response (color: more /s/ = yellow, more /ʃ/ = purple) for the interaction of continuum step (x-axis) and relative strength of the indicator of masculine stereotypes (y-axis). The three indicators of masculine toughness are represented individually by panel (left = toughness endorsement, center = talker toughness, right = face toughness).

stereotypes of toughness exhibited significant variation; although the responses were skewed toward a greater rejection of masculine stereotypes, there was a substantial tail toward a greater endorsement of masculine stereotypes.

While neither ToughnessEndorsement nor TalkerToughness emerged as significant main effects in the model, a main effect was observed for FaceToughness ($z = 2.28, p < 0.05$), with higher toughness ratings of the face generally predicting more /ʃ/ responses regardless of Step, i.e. regardless of how /ʃ/-like the stimulus is. Contrariwise, when taking Step into account, increased predictions of masculine stereotypes tend to predict fewer /ʃ/ responses. The interaction of Step with each of these three indicators is illustrated in Figure 4.8 using heatmaps. Heatmaps allow for the presentation of two continuous variables, here continuum step and toughness indicators, by transferring the response variable from the y-axis to the z-axis, represented two-dimensionally by hue. In these maps, a darker, purpler cell indicates a stronger prediction of an /ʃ/ response, while a lighter, yellower response indicates a stronger prediction of an /s/ response. If Step (on the x-axis)

is a strong predictor of categorization, this would be represented visually by a horizontal gradation from yellow to purple. If an indicator of masculine toughness is a strong predictor of categorization, this would be represented by a vertical gradation. If both factors interact to predict categorization, this would be represented by a diagonal gradation.

First looking at ToughnessEndorsement in the left-hand panel, while the horizontal gradation is most prominent, a gradation can also be observed vertically, suggesting a significant interaction of Step and ToughnessEndorsement. This is supported by the model ($z = -4.07, p < 0.001$), with listeners who more strongly endorse masculine stereotypes giving more /ʃ/ responses across the board, illustrating a less categorical perception of /SCr/ clusters. Turning to TalkerToughness in the center panel, only horizontal gradations are discernible, suggesting that a voice rated as more or less masculine is not influencing sibilant categorization. This is confirmed by the model ($z = -0.86, p = 0.39$), as the interaction of TalkerToughness and Step does not reach the threshold of significance, although it is worth noting that the negative direction of the effect trends in the same direction as ToughnessEndorsement. Finally turning to FaceToughness in the right-hand panel, again the horizontal gradations are most prominent, but, in contrast to TalkerToughness, a very subtle vertical gradation can be observed. This is confirmed by the model ($z = -2.24, p < 0.05$), with faces rated as more tough predicting more /ʃ/ responses across the board and contributing to a less categorical perception of sibilants in these clusters. The parallel findings of Step with ToughnessEndorsement and FaceToughness, as well as the trend of TalkerToughness that does not reach the significance threshold, suggest that an increased indication of masculine toughness serves to decrease the categoricity of sibilant perception in /SCr/ clusters, potentially speaking to the socio-indexicality of the onset sibilant, regardless of the cluster identity.

Additionally, ToughnessEndorsement also emerged as significant in its interaction with Cluster and Step ($z = -1.97, p < 0.05$). Again using a heatmap to visualize inter-

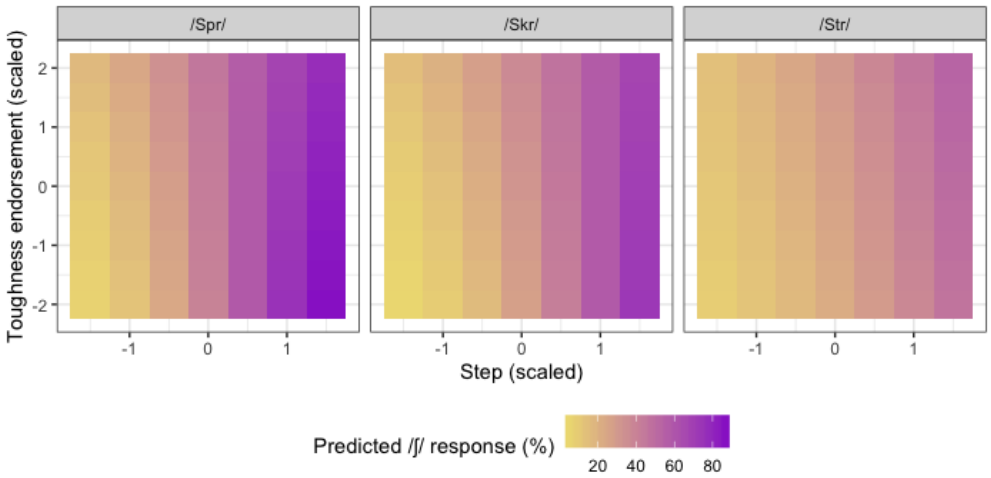


Figure 4.9: Predicted /ʃ/ response (color: more /s/ = yellow, more /ʃ/ = purple) for the interaction of continuum step (x-axis), toughness endorsement (y-axis), and cluster (panel: left = /Spr/, middle = /Skr/, right = /Str/). Clusters are ordered from right to left from most categorical to least.

actions between continuous variables, Figure 4.9 illustrates this interaction with each panel displaying a different cluster. The most salient difference between the panels is the relative difference in the amount of purple: A more pronounced yellow-to-purple horizontal gradient is observed in /Spr/ clusters in the left-hand panel; little purple at all is observed in /Str/ clusters in the right-hand panel; and /Skr/ in the middle panel falls somewhere in between. This horizontal gradation, however, is another representation of the interaction of Cluster and Step, represented as response curves in Figure 4.6. The three-way interaction of Step, Cluster, and ToughnessEndorsement is represented in the difference in a diagonal gradation between the panels, which is noticeably more pronounced in /Spr/ and /Skr/ clusters than in /Str/ clusters. After accounting for the observation that listeners in general are less categorical in /Str/ clusters than /Spr/ and /Skr/ clusters and the observation that listeners who more strongly endorse masculine stereotypes are less categorical than their less stereotype-focused peers, the interaction of ToughnessEndorsement, Cluster, and Step demonstrates that listeners who more strongly endorse masculine stereotypes exhibit an additional decrease in categoricity in /Spr/ and /Skr/ clusters relative to /Str/ clusters

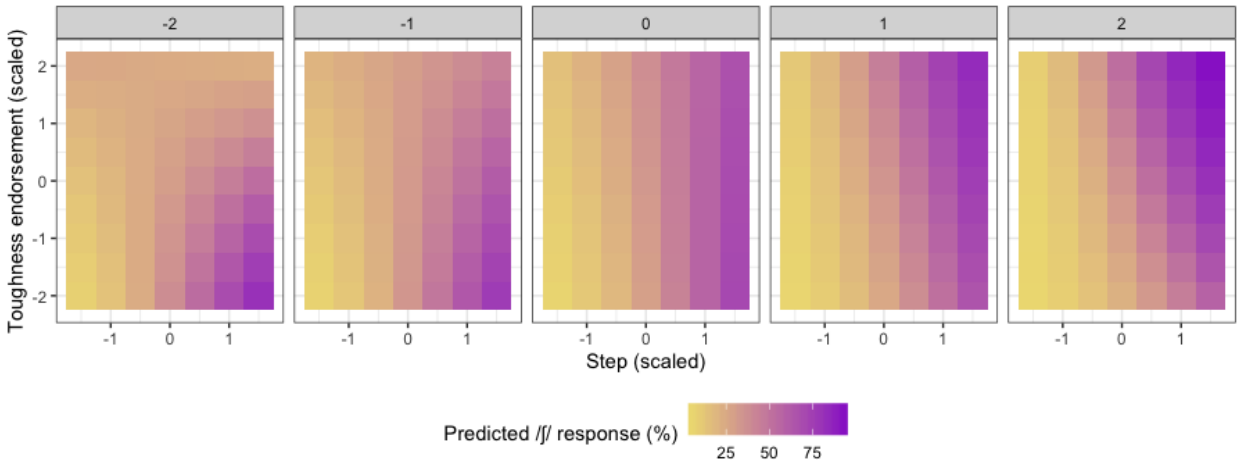


Figure 4.10: Predicted /ʃ/ response (color: more /s/ = yellow, more /ʃ/ = purple) for the interaction of continuum step (x-axis), toughness endorsement (y-axis), and age (panel: left = younger, right = older). Age is log-transformed and binned for ease of visualization but analyzed continuously.

compared to listeners who more strongly reject those stereotypes. This observation may suggest that the potential socio-indexicality of preconsonantal sibilants is more strongly linked to /Spr/ and /Skr/ than /Str/ clusters. In this way, listeners who attune to performances of masculine toughness are accounting for sibilant variation by attributing it to those performances, and they do so in /Spr/ and /Skr/ clusters where /s/-retraction is less expected.

Finally, after separately examining apparent time trends through listener age and socio-indexical meaning through indicators of masculine stereotypes, Figure 4.10 illustrates the model predictions for the interaction of ToughnessEndorsement, Step, and Age. In the leftmost panel (-2), a yellow-to-purple diagonal gradient can be observed from the top-left to the bottom-right, suggesting that the youngest listeners who most reject masculine stereotypes of toughness are more categorical in their perception of sibilants than similarly young listeners who more strongly endorse such stereotypes. As the participants approach the mean age, the predicted difference between those who endorse and reject masculine stereotypes dissipates, illustrated in the centermost panel (0) by the horizontal rather than diagonal gradation in /ʃ/ response. For the oldest listeners in the rightmost panel (2),

a distinct yellow-to-purple diagonal gradation can again be observed, but this time from the bottom-left to the top-right. This pattern suggests that the oldest listeners who more strongly endorse the masculine stereotypes are more categorical than similarly older listeners who reject those stereotypes, a pattern opposite to the youngest speakers in the leftmost panel. This observation is supported by the model, with the interaction of ToughnessEndorsement, Step, and Age ($z = 4.85, p < 0.001$) predicting more /ʃ/ responses at higher steps for older participants who more strongly endorse masculine stereotypes of toughness. This observation may suggest an apparent time change in the socio-indexicality of preconsonantal sibilants: Younger speakers, who by nature of their endorsements of masculine stereotypes are more attuned to performances of masculinity, may attribute increased retraction in /SCr/ clusters to such performances of masculinity.

4.5 Discussion

In the previous chapter, I demonstrate that listeners have detailed phonological knowledge about sibilant production and make use of the knowledge in real time. In the present experiment, I build off these findings to examine the categorization of potentially ambiguous sibilants in order to corroborate the findings of Experiment I: Cue Integration, examine potential time trends in /s/-retraction perception, and probe the potential socio-indexical meaning of the phenomenon.

First and foremost, this experiment asks if listeners attend to phonological environment when categorizing preconsonantal sibilants. Specifically, I ask if listeners make adjustments for their increased expectation of coarticulatory /s/-retraction in /str/ clusters compared to /spr/ and /skr/ clusters; that is, do listeners compensate for retraction and/or are some clusters perceived more categorically than others? The findings of this experiment confirm Hypothesis 1, suggesting that listeners do attend to phonological context, as they perceive sibilants less categorically in /Str/ clusters compared to /Spr/ and /Skr/ clusters. In fact,

even at the final step of the continuum, in which the sibilant was composed of 95% prevocalic /ʃ/, listeners still respond by categorizing the word as beginning with /s/ 50% of the time. However, listeners continue to perceive /s/ even at the final steps in /Spr/ and /Skr/ clusters as well, around 25% of the time, which is significantly less than in /Str/ clusters, but still much more often than would be expected in a categorization task for prevocalic sibilants. This result is a consequence of the phonotactic restrictions of English, by which /ʃ/ is illicit preceding consonants other than /r/, eliminating the phonological contrast between /s/ and /ʃ/ preceding consonants. It was these phonotactic restrictions that necessitated the use of nonce words to begin with in order to allow listeners to potentially make a contrast between the sibilants in these environments. Thus, it is not unexpected that listeners continue to hear /s/ at even the most /ʃ/-like steps, since this is phonologically more licit and orthographically more acceptable.

The difference observed in the categorization responses between the three consonant clusters, however, cannot be explained by phonotactics (or orthography) alone, as the same lack of contrast exists equally for /Str/, /Spr/, and /Skr/ clusters. Rather, the decreased categoricity observed in /Str/ compared to the other clusters is due to the increased expectation of coarticulatory retraction. Whether this constitutes compensation for coarticulation is less clear, as the adjustments observed for preconsonantal sibilants is manifested very differently than for prevocalic sibilants or other coarticulatory phenomena in which there are relevant phonological contrasts. Specifically, rather than the canonical shifted s-curve observed for other contrasts, including prevocalic /s/-/ʃ/, adjustments in /SCr/ clusters are manifested by a decreased categoricity, and thus a general flattening of the response curve, especially at higher, i.e. more /ʃ/-like, steps, as illustrated in Figure 4.6. Regardless of whether the decreased categoricity is truly perceptual compensation for coarticulation or not, these context-specific adjustments illustrate that listeners are accounting for phonological context and their expectations for coarticulation in the perception of preconsonantal

sibilants.

Secondly, with this experiment I ask if the apparent time trends that have been reported in the production of /s/-retraction (Gylfadottir, 2015; Smith et al., 2019; Wilbanks, 2017) are mirrored in the perception of these clusters. Specifically, I ask if younger speakers, by nature of their potentially greater familiarity with /s/-retraction, are more likely to account for it in their perception by compensating/perceiving these clusters less categorically than older listeners. The findings of this experiment support Hypothesis 2, as younger listeners are generally less categorical than older listeners in the perception of the onset sibilant in all /SCr/ clusters. However, as the interaction of listener age, cluster identity, and continuum step did not reach the significance threshold, the results of this experiment do not find that younger listeners are more or less categorical than older listeners in their perception of /Str/ clusters relative to /Spr/ and /Skr/ clusters. This result suggests that while younger listeners are less categorical across the board than older listeners, the relative difference in categoricity between the clusters has remained constant in apparent time. That is, just as younger listeners as less categorical in their perception of /Str/ clusters than older listeners, so too are they less categorical in /Spr/ and /Skr/ clusters. This observation has potential implications for the trajectory of the sound change, as, if perception and production of a sound change progress side by side, then as listeners increasingly perceive /Spr/ and /Skr/ clusters less categorically, so too are they expected to produce greater degrees of /s/-retraction in those clusters. This suggests that /skr/ and /spr/ clusters may be the next loci for /s/-retraction, which is tested in Experiment III: Convergence (Chapter 5).

Finally, this categorization task additionally asks how different indicators of masculine stereotypes influence the categorization of the onset sibilants in /SCr/ clusters in order to better understand the potential socio-indexicality of sibilant variation in these environments. This question is motivated by not only the evaluation of retracted prevocalic sibilants as more masculine (Campbell-Kibler, 2011a), the attunement to masculine stereotypes in the

categorization of prevocalic sibilants (Strand, 1999), and the online meta-commentary around /s/-retraction linking it to masculinity and toughness (Section 2.6.3), but also the contrasting findings of Phillips (2018) that a more retracted /s/ in consonant clusters is not robustly evaluated as more masculine than less retracted /s/. The findings of this experiment support Hypothesis 3 in part, suggesting that indicators of masculine toughness do play a role in the categorization of sibilants in these clusters, with most, but not all, effects predicting more /s/ responses. This suggests that listeners, especially those who may be more attuned to subtle performances of masculinity, attribute potential retraction to a performance of masculinity and account for that performance to recover the intended /s/ target.

However, while listeners are on average less categorical in /Str/ clusters than /Spr/ and /Skr/ clusters due to their familiarity with coarticulation in /str/ clusters, they are less categorical in /Spr/ and /Skr/ clusters than /Str/ clusters when accounting for potential performances of masculine toughness. This observation suggests that a retracted /s/ in /str/ clusters, precisely where /s/-retraction is most expected, less strongly indexes masculine toughness than a retracted /s/ in /spr/ or /skr/ clusters, where /s/-retraction is less expected. Thus, /s/-retraction, defined narrowly as an ongoing sound change limited to /str/ clusters, appears not to be primarily about performing masculinity and toughness. This finding corroborates the results of a social evaluation task from Phillips (2018), in which a stereotypically gay talker was evaluated as more heterosexual with a retracted /s/ in /skr/ and /spr/ but not /str/ clusters, but contrasts with the online meta-commentary linking /s/-retraction to masculine toughness. However, these observations from the online meta-commentary were limited to the lexical items *straight* and *strong*, suggesting that any potential indexing of masculine toughness may be linked to these lexical items rather than the phonological environment /str/ in general.

The findings of this experiment, however, should not be interpreted as implying that /s/-retraction is a sound change that does not do socio-indexical work. Rather, this experiment

simply finds that retraction in /str/ clusters does not robustly index masculine toughness, as defined by listeners' ratings of the faces and voices provided, and their own relative endorsement of masculine stereotypes of toughness, as defined by their responses to survey questions. Future research is necessary to better understand precisely how listeners perceive variation and innovation in these environments and what meaning they assign such variation beyond what is expected from research on prevocalic sibilants.

CHAPTER 5

EXPERIMENT III: CONVERGENCE

In the first two experiments of this dissertation (Experiment I: Cue Integration & Experiment II: Categorization), I focus on speech perception, finding that listeners have detailed knowledge (phonological, social, etc.) about /s/-retraction and use that knowledge in real time to make predictions and fine-tune their perceptual strategies. Experiment I: Cue Integration in particular clearly demonstrates that retraction is beneficial to the listener; it is a perceptually useful cue in distinguishing /sC/ and /sCr/ clusters and is available to listeners before the ultimate presence of the disambiguating /r/.

In the present chapter, I turn to production, or more accurately, the production-perception link, asking whether individuals will converge toward manipulated degrees of /s/-retraction. Convergence is the process by which a speaker begins to sound more like their interlocutor over the course of a conversation. In the laboratory, this is the process by which participants begin to sound more like a model talker over the course of an experiment. The first aim of this experiment is to test the hypothesis of a convergence path to sound change, whereby the shifts that occur over the course of a conversation persist and accumulate, eventually snowballing to lead to sound change (Auer & Hinsken, 2005; Nguyen & Delvaux, 2015; Trudgill, 1986, i.a.). With this experiment, I ask whether we can accelerate the propagation of the sound change, at least temporarily, in the laboratory. A sound change can be said to be accelerated through convergence if participants consistently converge toward the more advanced variants of the sound change. The second aim of the experiment is to better understand the potential constraints on convergence. These include phonological constraints (Will speakers will only imitate speech that enhances but does not diminish a phonological contrast?), coarticulatory constraints, (Will speakers only imitate speech that increases coarticulation and reduces articulatory 'effort'?), and social constraints (Will speakers only imitate speech from interlocutors with whom they identify on a relevant trait?).

In this chapter, I first present an introduction to the research on convergence (Section 5.1), including studies conducted both in and out of the laboratory and highlighting proposals for how convergence may help propagate sound change. I then provide a general overview to the study (Section 5.2) and introduce the materials, procedure, and hypotheses of this experiment (Section 5.3). Next, I present the results of the experiment (Section 5.4), a discussion of the findings (Section 5.5), and the implication this study holds for our understanding of phonetic convergence and its potential role in sound change actuation or propagation.

5.1 An introduction to convergence

Phonetic convergence is the process by which a speaker acquires some of the phonetic characteristics of their interlocutor. This process is also frequently called *imitation* or *accommodation*, including in many of the studies cited in this section (I discuss the selection of the term *convergence* for this dissertation in Section 1.1). At its core, phonetic convergence requires an integration of an individual's speech production and perception systems, as the lister-turn-speaker makes adjustments to their production as a result of what they perceive. Convergence has been observed in the laboratory in cooperative, socially-meaningful, and largely spontaneous conversations (Kim et al., 2011; Natale, 1975a,b; Pardo, 2006, 2009; Pardo et al., 2010, 2012, i.a.), as well as simple, shadowing tasks largely stripped of their social context (Babel, 2009, 2010; Babel & Bulatov, 2011; Babel, 2012; Delvaux & Soquet, 2007; Dufour & Nguyen, 2013; Goldinger, 1997, 1998; Goldinger & Azuma, 2003; Kraljic et al., 2008; Mitterer & Ernestus, 2008; Mitterer & Müsseler, 2013; Namy et al., 2002; Nielsen, 2011; Pardo et al., 2012; Pinget, 2015; Yu et al., 2013; Zellou et al., 2016, 2017, i.a.). Furthermore, outside of the laboratory, shifts have been observed in what is assumed to be the results of long-term interactions with individuals (Chambers, 1992; Harrington et al., 2000a,b; Harrington, 2006; Harrington et al., 2019; Payne, 1980; Sonderegger, 2012;

Sonderegger et al., 2017; Trudgill, 1981, 1986).

Much of the work on convergence has asked whether convergence is an automatic or conscious process, or whether it's somewhere in between. Some researchers have suggested that it's an automatic, phonetic process and thus unavoidable and ubiquitous (e.g. Goldinger, 1998; Pickering & Garrod, 2004; Trudgill, 2008). Other work has found that convergence is mediated by social factors and thus may be a conscious stylistic choice (e.g. Giles et al., 1991; Namy et al., 2002; Pardo, 2006; Shepard et al., 2001). Additionally, recent work has found that abstract, phonological knowledge plays a potential role in conditioning convergence (e.g. Nielsen, 2011; Zellou et al., 2016). In an extensive investigation of these various factors, Babel (2009) proposes that phonetic convergence is not automatic in terms of when and where it occurs, as it is mediated by both phonological and social factors, but that it is automatic in that speakers do not appear to making conscious, explicit choices to imitate, but rather are influenced by their subconscious implicit biases.

Phonetic convergence has been observed for a variety of different acoustic cues, from general prosodic parameters like overall intensity (Natale, 1975a), fundamental frequency (Babel & Bulatov, 2011; Gregory et al., 1993), and speaking rate (Webb, 1970) to phone-specific spectral and temporal cues, including vowel formants (Babel, 2010, 2012; Pardo et al., 2012), voice onset time (Nielsen, 2008, 2011; Pinget, 2015), fricative voicing (Pinget, 2015), and coarticulatory vowel nasalization (Zellou et al., 2016, 2017). Most notably for this dissertation, Kraljic et al. (2008) demonstrates convergence for sibilant spectral means. Many other studies of convergence approach convergence holistically and impressionistically, relying on listeners' judgments of similarity rather than analyses of the acoustic signal (Goldinger, 1997, 1998; Goldinger & Azuma, 2003; Pardo, 2006, 2009). This approach captures convergence broadly, as individuals may converge on different attributes or on multiple attributes simultaneously (Pardo et al., 2013).

While this dissertation is concerned with phonetic shifts, it is also worth noting that

convergence/imitation is not limited to phonetic factors in speech, as it has been observed for lexical (e.g. Brennan & Clark, 1996; Garrod & Anderson, 1987), syntactic (e.g. Bock, 1986; Branigan et al., 2000), semantic (e.g. Garrod & Anderson, 1987), and higher-order discourse factors, like spatial reference systems (e.g. Schober & Clark, 1989). Furthermore, convergence/imitation is not limited to language but extends to other behaviors, with facial expressions and yawning as perhaps the best known-examples (Bargh & Williams, 2006; Chartrand & Bargh, 1999; Dijksterhuis & Bargh, 2001; Meltzoff & Moore, 1977; Provine, 1989, i.a.). Nor is imitation limited to humans, with well-known examples of animals reproducing human vocalizations like the African Grey parrot (Pepperberg, 1981, 2007). However, while humans share the basic perceptual mechanisms necessary for imitation of other animals, humans appear to differ from other animals in that mental representations and other knowledge, like social information, mediate and inhibit the perception-production link (Dijksterhuis & Bargh, 2001). In such a proposal, social information inhibits, rather than facilitates or encourages, convergence, not the other way around.

Phonetic convergence, or the act of imitating speech in general, has also been proposed to play a critical role in language evolution, language acquisition, and language change. The discovery of the mirror system in the F5 area of the premotor cortex of macaques, which corresponds with Broca's area in the human brain, was found to be connected to perceiving and producing grasping actions. These findings have been interpreted to suggest that the development of dexterity as an imitative act gave rise to imitation of manual gestures and sounds, eventually playing a critical role in the evolution of the human faculty for language (Arbib, 2002, 2005). Similarly, other work has proposed that human language developed from the imitation of facial expression and vocalizations of Homo erectus (Studdert-Kennedy, 2000, 2005). In both of these accounts, imitation is not simple mimicry, but rather involves in the parsing of the perceived behavior into interpretable and categorizable components, providing the building blocks for structured language today. Synchronically, phonetic convergence,

or the act of imitating speech in general, has been proposed to play a critical role in language development (Charman, 2006; Masur & Eichorst, 2002), with children as young as 2-6 months-old (Kuhl & Meltzoff, 1996; Gratier & Devouche, 2011) exhibiting convergence and second language acquisition, like the acquisition of non-native speech sounds (Reiterer et al., 2013). Furthermore, phonetic convergence has been proposed to be a mechanism for individual longitudinal changes, like dialect acquisition (Trudgill, 1981, 1986), in which the shifts in a conversation are hypothesized persist well past the interaction. Moreover, if these shifts spread between individuals and communities, convergence has been hypothesized to explain language changes across a speech community (Auer & Hinskens, 1996; Auer & Hinsken, 2005; Nguyen & Delvaux, 2015; Nieldielzki & Giles, 1996; Trudgill, 1981, 1986). These hypotheses are explored more in the following section to ask how convergence can lead to various language change, like /s/-retraction.

In the remainder of this section, I highlight some of the research on convergence that motivates and situates the present experiment. I first examine the link between convergence and sound change through studies of convergence conducted outside the laboratory (Section 5.1.1). I move on to convergence in the laboratory (Section 5.1.2), examining the research focusing on the effects of coarticulation, sound change, and socio-indexicality. Finally, I situate discussions of convergence in prominent phonological theories, including exemplar theory (Section 5.1.3) and Communication Accommodation Theory (Section 5.1.4).

5.1.1 Convergence outside the laboratory

This present experiment examines convergence through the lens of sound change to ask if the small, conversational shifts can potentially play a role in the larger, longitudinal changes that make up sound change. As mentioned in the previous section, convergence has been hypothesized to be a mechanism for these changes, especially including cases of dialect contact and dialect leveling (Auer & Hinskens, 1996; Auer & Hinsken, 2005; Nguyen

& Delvaux, 2015; Nieldielzki & Giles, 1996; Trudgill, 1981, 1986). These proposals build off some of the earliest studies in dialectology, stemming from work in the late nineteenth century (e.g Paul, 1870, 5th edn. 1920). In this section, I examine a convergence path to sound through studies of dialect contact and longitudinal corpus examinations of intraspeaker variation.

As outlined by Auer & Hinsken (2005), a convergence path to sound change (in their words, 'change-by-accomodation'), involves three components:

1. Change in a conversation: A speaker adopts a new feature from their interlocutor.

2. Change in the individual: The convergence is lasting; it permanently affects the speaker.

3. Change in a community: The innovation spreads throughout the speaker's community.

Such a proposal creates an implicational hierarchy between these components where the short-term conversational shifts must first take place (step one) in order for such changes to be lasting within the individual (step two) in order for those changes to spread across a community (step three). Much of the evidence examined for such a proposal focuses on dialect convergence and leveling, with individuals more likely to converge toward the standard dialect during a conversation (Hinskens, 1996), long-term evidence of individuals converging toward their new dialect group (Trudgill, 1981; Hinskens, 1996), and longitudinal evidence of dialect leveling across different speech communities (Gilles, 1998).

In one such study, Trudgill (1981, 1986, 2008) proposes that an individual in contact with a novel dialect group will exhibit convergence toward salient dialectal markers, where markers are understood to carry social meaning and are above the level of consciousness. Trudgill suggests that this convergence emerges out of a desire to be understood with a new group and begins at lexical differences before moving onto phonetic differences. Crucially, Trudgill argues that although phonetic differences may be the target for dialect convergence if they're

sufficiently stigmatized or phonetically dissimilar from the speaker's own dialect, individuals exposed to dialect variation will not converge toward different phonological systems. Trudgill (1986, p. 58, emphasis in original) argues that "[s]peakers' motivation, moreover, is *phonetic* rather than phonological": their purpose is to make individual words sound the same as when pronounced by speakers of the target variety". In support of these claims, Trudgill (1981) examines the speech of children from different dialect areas who move to Norwich, England, finding that they converge toward phonetic differences but maintain their original phonological systems. For example, Norwich maintains a contrast between the historic /uː/ (*nose*) and /ʌu/ (*knows*) vowels, unlike most dialects of English; Received Pronunciation collapses both these vowels classes to /əʊ/ and standard American to /oʊ/. Children new to the Norwich dialect area are unable to consistently distinguish these vowel classes, failing to acquire the new dialect phonological contrasts.

Similar findings have been reported by Payne (1980) and Chambers (1992). Payne (1980) examines dialect convergence for children who move to King of Prussia, Pennsylvania. All of the children were observed to exhibit phonetic convergence toward their new dialect group but experienced difficulty in acquiring new phonological rules, like the tensing of /æ/ before nasals. Similarly, Chambers found the same result for Canadian children moving to New England: They produce the phonetic forms of their new dialect area but often not the new phonological rules. Later in his career, Trudgill (2008) proposed that the shifts of the kind of observed in these and many other studies are the result of an automatic phonetic process and not mediated by sociolinguistic factors. While such claims remain contentious, the empirical results of these studies are definitive: Individuals, especially children, new to an area exhibit phonetic shifts toward the new dialect group. Furthermore, when individuals remain in contact with the new dialect group, it can be presumed that the effects of this convergence are persistent and potentially cumulative, eventually leading to an observation of change within the individual. These observations suggest that if dialect change observed on the

individual level is the result of cumulative convergence, sound change within a community may also be explained via a similar path.

While dialect leveling supports a convergence path to change, it does not explain how innovative forms arise or spread between individuals. That is, can the same process explain both novel innovation and dialect leveling? Given that innovation, like /s/-retraction, emerges and propagates through means other than dialect contact, this dissertation focuses on step one of the convergence path to sound change, asking if individuals converge toward the innovative form in short-term interactions. In order for these short-term shifts to lead to lasting change, conversational convergence must have lasting effects. Such lasting effects of conversational shifts can be glimpsed, albeit indirectly, through longitudinal studies of intraspeaker variation.

Using over thirty years of annual Christmas broadcasts, Harrington and colleagues (Harrington et al., 2000a,b; Harrington, 2006) investigate the speech of Queen Elizabeth II, finding that her vowel space from the 1980s appears to have shifted significantly from it where it was in the 1950s, and that as a whole it has moved toward the more standard Southern British vowel space. These results may suggest that the Queen's shifts in pronunciation are an instance of dialect convergence, potentially as a result of increased exposure to speakers of Southern British English. Looking at a shorter but more continuous timespan, Sonderegger and colleagues (Sonderegger, 2012; Sonderegger et al., 2017) examine VOT, coronal stop deletion, and vowel formant convergence on the television show *Big Brother*, in which contestants are confined to a single house without outside contact for three months. Sonderegger (2012) reports individual differences in the short- and medium-term shifts observed. For most individuals, short-term (i.e. day-to-day) shifts are observed, but for some individuals these short-term shifts appear to have trajectories suggesting an incremental approach to change. Specifically, the two individuals who spent the most time together and had a positive connection with one another as a result of their apparently genuine romantic relationship

that emerged during the filming, were observed to exhibit shifts on various phonetic cues. For these two speakers, their medium-term shifts appeared to be manifestations of incremental convergence, as their production approached one another in Bark transformed vowel formants and VOT. However, it remains unclear whether those medium-term shifts persisted after the filming ended. In an examination of oral arguments before the U.S. Supreme Court over the course of a year, Yu et al. (2015) find that all the Supreme Court justices exhibit significant variation in vowel-to-vowel coarticulation. However, for the eight individuals examined over a 205 day span, Yu et al. found no evidence for time trends, suggesting that, despite day-to-day variation, the coarticulatory effects are relatively stable over time. Finally, in an examination of a sample of individuals relatively isolated for several months at a research station in Antarctica, Harrington et al. (2019) found that individuals incrementally converged on vowel backness for /ɪː/, /u/, and /ju/ and innovated a change by which /ou/ was fronted that does not appear to be the result of convergence. These findings, although limited in scope, may be interpreted to suggest that accent development and sound change can emerge both as a result of convergence and for independent reasons.

5.1.2 Convergence in the laboratory

In the laboratory, convergence has been robustly examined with researchers utilizing a variety of tasks and focusing on a diverse set of cues and measurements, including holistic similarity judgments. Goldinger (1998) extracted single words from a shadowing task and placed them in AXB similarity task, in which listeners were asked to decide which utterance produced by the participant in the previous experiment (A or B) sounds more like the model talker (X). One of the tokens from the participant was taken from the reading task before hearing the model talker (A) and the second was from the shadowing task (B), in which the participant repeated after the model talker. Goldinger found that the utterance produced shadowing the model talker was perceived to sound more like the model talker. Moreover, the strength

of this effect was mediated by the amount exposure to the model voice and the gap between the model voice and the shadowing task. Consequently, Goldinger takes these results to suggest that convergence is mediated by episodic memory (more on that in the following section). Using the same methodology, Namy et al. (2002) additionally found a gender effect, with female participants converging to both male and female model speaker more than male participants.

Pardo (2006, 2009) demonstrated that convergence can be observed in the laboratory outside of a shadowing task when individuals are paired and participating in a meaningful, cooperative task, like the map task. In a map task (Anderson et al., 1991), one participant, the instruction giver, is given a hand-drawn map containing various landmarks with corresponding labels (e.g., 'wheat field' and 'bear cave') and the path taken, including starting and ending points. Their interlocutor, the instruction receiver, is given a similar map, though notably missing the route, starting and ending points, as well as minimally differing landmarks/labels. The instruction giver then guides the receiver through the map. The design of the map task encourages both participants to say the target words, represented in the landmarks, in a naturalistic and spontaneous fashion. Following the map task, Pardo (2006) extracted single words and embedded them in an AXB similarity task, where the X items came from the participant's conversational partner. Like Namy et al. (2002), Pardo found unexpected gender effects: While female instructor receivers were more likely to converge to their instructor giver, male instruction givers were more likely to converge to their instruction receiver.

Turning away from perceived similarity, Babel (2009) demonstrated convergence in the acoustic signal, with participants converging toward the vowel formants of the model talker. Babel found evidence for shifts in the first and second formants of some vowels, with /ɑ/ and /æ/ showing the strongest convergence effects in the first formant. Furthermore, Babel (2009) found evidence for an effect of the perceived race of the model talker, with white

speakers showing less convergence to black talkers. Babel (2010) expanded this work to examine nationality, examining convergence of New Zealand speakers shadowing an Australian talker. Again Babel (2010) found evidence for acoustic convergence for some of the vowels, and once again there was an effect of social factors, as New Zealanders more oriented toward Australia converged more toward the Australian vowels. Examining ongoing sound changes in Dutch, Pinget (2015) developed a creative covert shadowing task, in which the participant played a computer card game with two model talkers. Pinget found that Dutch speakers converged to bilabial stop devoicing and labiodental fricative devoicing, both in line with the changes in progress.

Three studies are of particular note to this dissertation, as they each find evidence for directionality and two of them examine the convergence of coarticulation. Firstly, Nielsen (2011) examined convergence in the laboratory toward manipulated degrees of VOT. Participants read a series of words containing word-initial stops both silently and out loud in isolation. The participants were then exposed to a model talker who read with either increased or decreased VOT. In the increased VOT condition, the model talker's VOT was extended by 40 ms, and in the decreased VOT condition, the model talker's VOT was decreased by 40 ms. Following exposure to the model talker, participants repeated the reading task. Nielsen examined the shifts that occurred between the pre- and post-exposure reading tasks and found that participants exhibited convergence in the increased VOT condition, but no change in the decreased VOT condition. These findings suggest that phonetic convergence is a selective, rather than automatic, process, and that the directionality in the observed convergence may be explained by linguistic factors like phonological contrast. Specifically, participants may have avoided convergence in the decreased VOT condition due to the potential loss or reduction of contrast between voiced and voiceless stops, but converged in the increased VOT condition as the phonological contrast was enhanced rather then reduced.

Similarly, Zellou et al. (2016) examined convergence of nasal coarticulation on preceding

vowels. Like this dissertation, the question at hand is whether individuals exhibit convergence to model talkers with an increased degree of anticipatory coarticulation. Participants first completed a pre-exposure word-naming task followed by an exposure phase lexical identification task. Following these tasks, participants completed a shadowing task in which they were asked to repeat CVN(C) words (e.g. *den* or *band*) with artificially increased or decreased degrees of vowel nasalization. In the increased nasality condition, the vowel was cross-spliced from an NVN sequence to have increased degrees of vowel nasalization; for example, in *den* the vowel was cross-spliced from *nen*. In the decreased nasality condition, the vowel was cross-spliced from a CVC sequence, for example *dead* for *den*. Finally, participants completed a post-exposure word-naming task identical to the pre-exposure task. Zellou et al. found convergence to the model speaker who exhibited increased nasality but acoustic divergence from the model speaker who exhibited decreased nasality (although possibly convergence relative to the speakers' phonological system). Like Nielsen (2011), these results suggest that convergence is a selective process, although with Zellou et al. finding divergence in one condition and convergence in another. The directionality effect observed here may be due to the phonetic naturalness of the stimuli, such that more coarticulation is more phonetically natural, or due to alignment to a sound change in progress, as proposed by Zellou & Tamminga (2014).

In their examination of how listeners learn dialectal (i.e. /s/-retraction in /str/ words like *industry* or *artistry*) or idiolectal (i.e. /s/-retraction in /s/ words like *dinosaur* and *medicine*) patterns of /s/ variation, Kraljic et al. (2008) also ask whether individuals exhibit convergence toward the model talker, and if they do, whether explicit instructions are necessary in order to ensure that convergence is observed. In the first experiment, following exposure to the model talker, participants were explicitly asked to imitate the accent of the model talker in a shadowing task. Regardless of whether they were exposed to idiolectal or dialectal variants, participants from both conditions converged toward the model talker in

terms of spectral mean in /str/ clusters. For the majority of speakers, convergence was realized by increased retraction (i.e. lower spectral mean). But even the speakers who exhibited more retraction (i.e. lower spectral mean) than the model talker in the pre-test exhibited convergence during the shadowing task, raising their spectral mean in these clusters. These findings suggest that directionality does not play a role in the convergence of /s/-retraction, as speakers converge toward both increased and decreased retraction depending on their baseline. However, it is crucial to note that degree of retraction was not manipulated; participants differed in whether they had previous exposure to it before the shadowing task. Furthermore, speakers were explicitly asked to imitate the model talker with a strong, stereotypical accent. In the second experiment, individuals were not asked to imitate the model talker, but rather completed a pre- and post-exposure story completion with the mode talker. Regardless of their baseline degree of retraction and whether they were exposed to /str/ clusters in the category identification task, participants did not exhibit significant shifts toward the model talker. It is unclear whether individual variation was observed as only group analyses were presented. These findings, like Nielsen (2011) and Zellou & Tamminga (2014), strongly suggest that convergence is a selective process, as it occurs when speakers are asked explicitly to imitate. However, when participants are given no instruction, and rather we compare pre- and post-exposure measurements, no convergence is observed. These findings also indicate that when it is observed, convergence of /s/-retraction is not characterized by the directionality effects observed for vowel coarticulatory nasalization and VOT.

Taken together, the findings of these three studies make potentially contradictory hypotheses about the role of directionality in /s/-retraction convergence. Firstly, following Kraljic et al. (2008) if spontaneous convergence is observed, we expect no difference between increased or decreased degrees of /s/-retraction. However, the findings of Kraljic et al. (2008) also make contradictory predictions about whether convergence toward manipulated degrees of retraction would occur at all in an unprompted task. Secondly, following Zellou

et al. (2016), if spontaneous convergence is observed, speakers may be more likely to converge toward increased retraction, as it is both more phonetically natural and aligned with a sound change in progress. Thirdly, following Nielsen (2011), if spontaneous convergence is observed, speakers may be more likely to converge in the decreased retraction condition, as it is potentially less ambiguous, maximizing the contrast between /s/ and /ʃ/. However, as English makes no contrast between /s/ and /ʃ/ in these environments, convergence toward increased retraction can be observed without decreasing the contrast between the sibilants.

5.1.3 Convergence and exemplar theory

Examinations of convergence have provided rich testing ground for phonological theory and have particularly been suggested to bolster exemplar models of phonology. Within exemplar theory, as put forward by Pierrehumbert (2001), each sound, word, voice, etc. is associated with multiple mental representations, or episodic traces, encoding various phonetic and social details, collectively creating an exemplar 'cloud'. When a word or voice is heard, all of the associated episodic traces are activated, with more frequent words and more familiar voices activating a greater number of episodic traces. Additionally, traces have been proposed to decrease in their relative strength, or 'decay', over time. Due to the incredibly variable nature of the speech signal, exact matches between the token heard and the episodic trace are not expected; rather, all the similar traces are activated, creating a 'generic echo' of the mean of the activated set (Goldinger, 1997, p. 46).

For studies of convergence, exemplar theory suggests that when speech is planned, the associated exemplar clouds with the relevant episodic traces are activated. Since the most recent attestations of the planned sound or word are the strongest, these sounds are predicted to have the greatest influence in planning the speech targets. Thus, immediately following exposure to a model talker, the speaker is predicted to have targets shifted toward the model talker, resulting in the convergence observed in the numerous studies cited in this section.

Firstly, Goldinger (1998) varied the time between the exposure to the model talker and the repetition, finding that extracted tokens are perceived to be more similar to the model talker when the lag between exposure and production is shorter. This finding supports exemplar models as more episodic traces are predicted to decay over time. Similarly, Goldinger also examined the amount of exposure the participant had to the model talker, finding greater perceived convergence for individuals with more exposure to the model talker. This finding supports exemplar models as increased exposure would increase the number of episodic traces from the model tallker. Finally, Goldinger found that the perceived convergence is greater for low frequency words than for high frequency words. This finding bolsters the exemplar model in which high frequency words are expected to have more episodic traces, thus reducing the relative weight of the traces associated with the model talker.

However, as it is not universally observed on all phonetic metrics, for all phonological units, in all social settings (e.g. Babel, 2010, 2012; Nielsen, 2011; Pardo et al., 2012; Sonderegger, 2012; Sonderegger et al., 2017; Zellou et al., 2016), convergence cannot solely be the consequence of episodic traces. Rather, other elements like social attitudes and linguistic factors, possibly including phonetic naturalness and phonological contrast, must be integrated into exemplar accounts in predicting the observed convergence.

5.1.4 Convergence and Communication Accommodation Theory

Convergence has also served as a rich testing ground for the social nature of language by linguists and social psychologists alike. Chief among them, Giles and colleagues (Giles, 1973; Giles et al., 1973; Giles & Smith, 1979; Giles et al., 1991; Shepard et al., 2001) developed the Communication Accommodation Theory (CAT) to explain the social behavior around convergence. At the center of CAT is the hypothesis that individuals make adjustments in their speech to mirror their social situation, selecting one of four speech strategies: convergence, divergence, maintenance, or complementarity. For convergence, speakers minimize linguistic

(e.g. phonetic) differences in order to minimize social differences to seek social approval and acceptance. Conversely, for divergence, speakers enhance linguistic differences to enhance those same social differences, expressing opposition or disagreement.

The Communication Accommodation Theory has been adopted by some linguists to explain why and where convergence is observed in the real world and in the laboratory. For example, Trudgill (1986) initially embraced CAT as a model for understanding dialect convergence and change. In the laboratory, Namy et al. (2002) suggest that an appeal to social dynamics may help explain the observed gender differences observed in convergence. Similarly, Babel (2009) appeals to the social aspect of CAT to account for the differences in convergence with black and white model talkers which suggest that convergence is mediated by the speakers attitudes toward their interlocutor.

Just as exemplar models alone cannot explain exactly when and where convergence is observed, CAT cannot explain which phonetic variables are targets of convergence, as variables both below and above the level of consciousness can undergo convergence. Rather, CAT provides just another piece to the puzzle. For this dissertation, CAT predicts that differences in convergence (or divergence) patterns may be indicative of individuals' relative familiarity with /s/-retraction and/or the social associations it carries for them. This predicts convergence toward increased retraction for individuals wanting to express solidarity with the communities they associate with /s/-retraction, but divergence from increased retraction for individuals wanting to distance themselves from the (possibly different) communities they associate with /s/-retraction.

5.2 Study overview

Experiment III: Convergence asks if speakers converge toward their interlocutor for a sound change in progress. Specifically, I ask if participants shift their production of onset sibilants in /sCr/ clusters in response to the manipulated tokens produced by a model talker. In

this experiment, participants are not explicitly asked to imitate or shadow the model talker; rather, I compare their recordings prior to and following exposure to the model talker. The procedure roughly follows Nielsen (2011) with a pre-exposure baseline reading phase, an exposure phase in which participants passively listen, and a post-exposure reading phase.

Like Experiment I: Cue Integration (Chapter 3), there are three conditions for this experiment: decreased, increased, and hyper-increased retraction. The experiment employs a between-subject design, meaning that each participant was assigned to one condition and only heard stimuli with one manipulated degree of retraction. A between-subject design not only provides participants with exposure to an interlocutor who consistently produces retraction, such that an individual token cannot be interpreted as aberrant, but also ensures that all stimuli have the same potential effect on the participants' production, since pre- and post-exposure values are compared rather than using a shadowing task. Additionally, the different retraction conditions allow for an examination not just between the different places of articulation, but also between the different degrees of retraction within the same cluster.

In Section 5.3, I outline the methods and materials including the stimuli (5.3.1), participants and procedure (5.3.2), and analysis and hypotheses (5.3.4). In Section 5.4, I present the results of this experiment, and in Section 5.5, I move onto a discussion of the findings.

5.3 Methods & materials

5.3.1 Stimuli

Like in Experiment I: Cue Integration (Chapter 3), the auditory stimuli for this experiment are designed to target /s/ in various onset clusters containing manipulated degrees of /s/-retraction that can serve to provide the potential basis for convergence. Stop-initial stimuli containing manipulated VOT are also included in order to serve as fillers against the all sibilant-initial targets and to test the validity of the experimental design, as Nielsen (2011)

	Reading and listening			Reading only		
	word	IPA	$SUBTL_{WF}$	word	IPA	$SUBTL_{WF}$
/s/	sip	/sɪp/	5.10	sick	/sɪk/	165.43
	sit	/sɪt/	311.35	sift	/sɪft/	0.75
	sue	/su/	29.37	soon	/sun/	257.65
	suit	/sut/	68.61	soothe	/suð/	1.29
/sp/	spit	/spɪt/	19.35	spin	/spɪn/	14.63
	spoon	/spun/	7.61	spool	/spul/	0.51
/spr/	spritz	/sprɪts/	0.49	spring	/sprɪŋ/	31.31
	spruce	/sprus/	1.1	sprue	/spru/	0.00
/st/	sting	/stɪŋ/	7.02	stint	/stɪnt/	0.75
	stew	/stu/	6.43	stool	/stul/	3.51
/str/	string	/strɪŋ/	12.67	strip	/strɪp/	15.69
	strewn	/strun/	0.37	strudel	/struɾəl/	0.92
/sk/	skip	/skɪp/	21.1	skin	/skɪn/	44.04
	scoop	/skup/	5.67	scoot	/skut/	2.1847
/skr/	script	/skrɪpt/	19.61	scribble	/skrɪbəl/	0.63
	screw	/skru/	37.49	scrooge	/skrudʒ/	3.86
/sh/	ship	/ʃɪp/	98.88	shift	/ʃɪft/	22.82
	shit	/ʃɪt/	474.65	shin	/ʃɪn/	3.08
	shoe	/ʃu/	30.39	shoes	/ʃuz/	30.39
	shoot	/ʃut/	164.94	chute	/ʃut/	3.61

Table 5.1: Sibilant-initial wordlist for Experiment III: Convergence

observed robust convergence to increased VOT. The same auditory stimuli from Experiment I: Cue Integration were used here, although expanded to include more items than was feasible in the visual world paradigm. In addition to the stimuli from Experiment II that all contain /ɪ/ as the syllable nucleus, the present experiment also examines words with /u/. The onsets for the /u/-nucleic words were made following the same procedure described for /ɪ/-nucleic words. For details on stimuli creation and design, please see Section 3.3.1.

The primary focus of the present experiment is to first ask if individuals exhibit con-

vergence in /sCr/ clusters. Secondly, I ask whether differences in convergence patterns can be observed between the retraction conditions and/or the different clusters, i.e. the identity of the intervening stop in these /sCr/ clusters. To provide a reference point for each individual's production of /s/ in these clusters, prevocalic /s/ and /ʃ/ as well as /sC/ clusters were included. There were a total of 20 sibilant-initial auditory stimuli for the listening task which include an /ɪ/ and /u/ token for each cluster, plus two for prevocalic /s/ and /ʃ/. An additional 20 items were included for just the reading portion, yielding 40 items total. The full set of sibilant-initial stimuli for reading and listening tasks, including the /ɪ/-nuclear stimuli from Experiment I: Cue Integration and the /u/-nuclear words new to this study, are presented in 5.1, along with phonemic representations and lexical frequency values.

Like in Experiment I: Cue Integration, stop-initial stimuli were included as fillers since all the target stimuli were sibilant-initial. Stop-initial words were additionally selected, as individuals have been demonstrated to converge to increased VOT but not decreased VOT (Nielsen, 2011, for details see Section 5.1.2). The stop-initial class of words can thus serve not only mask the intent of the study, but also to provide a proof-of-design check for experiment. There were 24 stop-initial auditory stimuli for the listening task, an /ɪ/ and /u/ token for each cluster, with three places of articulation and simplex/complex onsets. An additional 24 items were included for just the reading portion, yielding 48 items total. The full set of stop-initial stimuli for the imitation task, including the /ɪ/-nuclear stimuli from Experiment I: Cue Integration and the /u/-nuclear words new to this study, are presented in 5.2. Additionally, this analysis assumes a yod-dropping dialect, such that *tune* is phonemically /tun/ not /tjun/ or /tʃun/.

	Reading and listening			Reading only		
	word	IPA	$SUBTL_{WF}$	word	IPA	$SUBTL_{WF}$
/p/	pig	/pɪg/	39.14	pick	/pɪk/	198.39
	pooch	/putʃ/	1.29	pool	/pul/	46.98
/pr/	prick	/prɪk/	14.12	prim	/prɪm/	0.37
	prune	/prun/	1.47	prove	/pruv/	70.39
/t/	tip	/tɪp/	27.63	tin	/tɪn/	8.65
	two	/tu/	1066.35	tune	/tun/	15.61
/tr/	trip	/trɪp/	82.39	trim	/trɪm/	4.27
	true	/tru/	253.35	truth	/truθ/	192.18
/k/	kit	/kɪt/	17.65	kick	/kɪk/	73.41
	coop	/kup/	10.35	coo	/ku/	0.69
/kr/	crypt	/krɪpt/	1.37	crimp	/krɪmp/	0.43
	crew	/kru/	47.53	crude	/krud/	3.04
/b/	big	/bɪg/	682.82	bin	/bɪn/	5.37
	boot	/but/	11.14	boom	/bum/	21.80
/br/	brick	/brɪk/	10.18	brim	/brɪm/	0.88
	brew	/bru/	2.51	bruise	/bruz/	3.24
/d/	dip	/dɪp/	7.96	din	/dɪn/	1.18
	dew	/du/	2.14	dune	/dun/	1.00
/dr/	drip	/drɪp/	5.12	drink	/drɪŋk/	247.39
	drew	/dru/	25.04	drool	/drul/	2.16
/g/	gift	/gɪft/	64.51	gill	/gɪl/	1.71
	goop	/gup/	0.69	goof	/guf/	2.22
/gr/	grip	/grɪp/	9.69	grill	/grɪl/	4.45
	group	/grup/	73.76	groove	/gruv/	4.16

Table 5.2: Stop-initial wordlist for Experiment III: Convergence

5.3.2 Participants & procedure

Participants

Seventy-five participants were recruited from the University of Chicago and the greater Chicago area. Initial participants were randomly assigned to either the increased or decreased retraction condition, as well as the increased or decreased VOT condition. Later recruits were automatically assigned to the hyper-increased retraction condition and either increased or decreased VOT condition. This yields 6 possible combinations of retraction and VOT conditions.

All participants were between the age of 18 and 22 (mean=20). Undergraduate students enrolled in introductory linguistics classes participated for course credit. Participants not participating for credit were paid at the rate of $20 per hour. All participants were self-reported native speakers of American English and grew up predominately in the United States. The geographic distribution of the participants is mapped in Figure 5.1, with notable gaps in much of New England, the plains, the Rockies, and the inland South. Most participants (49) self-identified as growing up in a suburban area, with fewer in urban (22) or rural (4) environments. A majority of the participants identified as straight/heterosexual (51) and a plurality of participants identified as white (36), with two or more races (15) or Asian (13) as the next most reported racial backgrounds.

Like Experiment I: Cue Integration (Chapter 3), participants completed an extensive post-task demographic survey, including the self-reported information provided above, as well as the series of sub-surveys described in the previous chapter to determine their relative measurements on a variety of social and cognitive scales that may influence the likelihood of convergence. On the Male Role Attitude Scale (MRAS, Pleck et al., 1993), participants in the present experiment had a mean composite score of 14.92 (s.d. 2.73, on a scale of 10 to 40) and a mean score of 3.78 (s.d. 1.34, on a scale of 2 to 8) on the toughness subscale, where higher

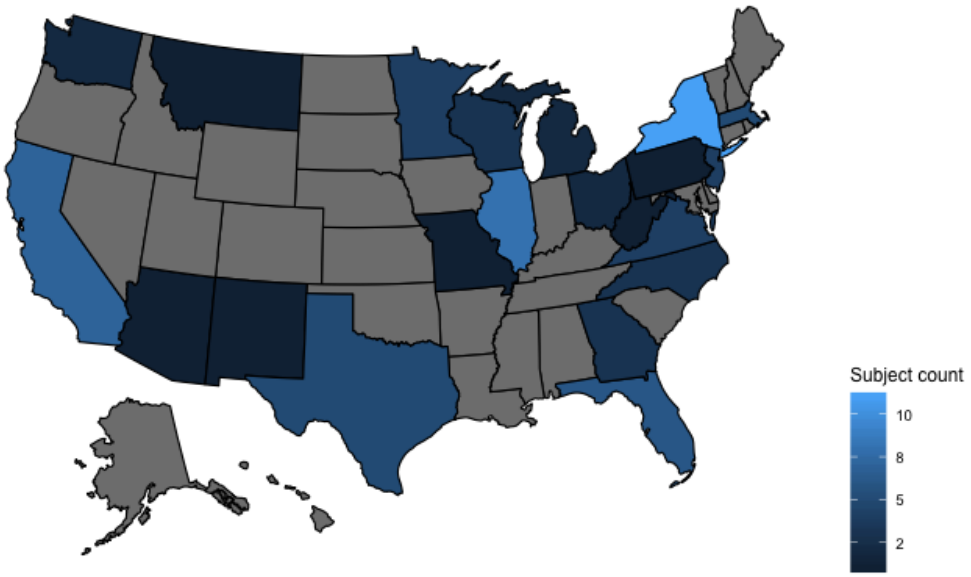


Figure 5.1: Geographic distribution of participants by self-reported state lived in the longest before age 18, with a lighter blue indicating a higher count of participants. Gray indicates no participants reporting living in that state.

scores indicate a stronger endorsement of traditional stereotypes. This demonstrates that participants still widely rejected masculine stereotypes in general, and masculine stereotypes of toughness specifically, at levels comparable to the previous experiment. The MRAS was included to account for listener variation in the attribution of /s/-retraction to masculinity. On the Empathy Quotient (Baron-Cohen & Wheelwright, 2004), the mean score for the present experiment was 16.82 (s.d. 8.64) on a scale of 0 (not at all empathetic) to 44 (very empathetic), suggesting that participants in the present experiment were less empathetic than the previous experiment. The Empathy Quotient was included to account for cognitive differences between listeners that may account for likelihood of exhibiting convergence, as individuals with poorer attention switching skills, often linked to other autistic traits, are less likely to exhibit convergence (Yu et al., 2013). On the The PROMIS Anxiety instrument (Cella et al., 2010), the mean score was 16.91 (s.d. 5.79) on a scale from 7 (not at all anxious) to 45 (very anxious). The Anxiety instrument was included as previous work has suggested that individuals who are more anxious are less likely to exhibit convergence (Lewandowski

& Jilka, 2019). Finally, on the openness dimension of the Big Five Inventory (John et al., 2008), the mean score was 28.57 (s.d. 4.82) on a scale from 8 (less open) to 40 (more open), comparable to the previous experiment, as individuals who are more open are more likely to exhibit convergence (Lewandowski & Jilka, 2019; Yu et al., 2013). It should be noted that these cognitive factors were included not as a primary focus of the present experiment, but in order to account for potential variation between participants, especially given the between-subject design of this study.

No participants included in the analysis reported a history of hearing loss, language and communication disorders, stroke, traumatic brain injury, or other medical or neurological conditions commonly associated with cognitive impairment. An additional twenty-one individuals participated in the present experiment but were excluded due to non-native status, reported language or neurological disorders, and/or recorder malfunction.

Procedure

The experiment was composed of four ordered blocks of three different tasks: (1) a silent reading task, (2) a pre-exposure reading task, (3) an exposure listening task, and (4) a post-exposure reading task. All participants completed all four blocks in the specified order. The experiment was presented on a computer monitor using PsychoPy (Peirce, 2007), with participants controlling the rate of the experiment by pressing the space bar to advance between items and blocks.

The study was conducted an isolated double-walled sound booth in the Phonology Laboratory at the University of Chicago.

Silent reading task Prior to recording participants' baseline measurements, participants were first asked to silently read the list of target and filler words (Tables 5.1 & 5.2), with explicit instructions to "read the words SILENTLY" and to "try not to mouth along as you

read". Each item was presented once in a random order. This warm-up task was included in order to reduce possible hyper-articulation, and thus possible avoidance of coarticulation, for low frequency words as found by previous research (Goldinger & Azuma, 2004; Nielsen, 2011).

Pre-exposure reading task Immediately preceding the exposure listening task, baseline measurements of participants' production of the target sounds were obtained prior to exposure to the mode talker. Participants were asked to read the same target and filler words from the silent reading task (Tables 5.1 & 5.2) but this time to "speak each word into the microphone as naturally as possible". Each item was repeated twice in a randomized order, yielding 196 trials (40 target items + 48 fillers $\times$ 2 = 196). Participants were fitted with a Shure SM10A head-mounted microphone and recorded on a Zoom H6 at a sampling rate of 48,000 Hz.

Exposure listening task Following the pre-exposure reading task, participants removed their head-mounted microphone and put on Sennheiser HD 555 circumaural headphones. The auditory stimuli (see Section 5.3.1 for details) were played binaurally at the rate of one word every two seconds. The auditory stimuli were paired with a colored circle that pulsed to the rate of the recordings, but otherwise no visual representation was present. Participants were instructed to listen to and "try to silently identify the word spoken to yourself". Each item was presented four times, yielding 392 trials (20 target items + 24 fillers $\times$ 4 = 176). Recall that only half of the items included in the pre- and post-exposure reading tasks were included as auditory stimuli in the listening task; this serves to ask if potential convergence is observed in the same phonological environments extended to new words.

Post-exposure reading task Immediately following the exposure listening task, participants removed their headphones and, to avoid interference in the convergence process from

the researcher, refit their own head-mounted microphones. I was then able to visually confirm through window in the sound booth that the microphone had been correctly refitted. Participants were then asked to again "speak [each word] into the microphone, as naturally as possible". Each item was repeated twice in a randomized order, yielding 196 trials (40 target items + 48 fillers × 2 = 196). The post-exposure reading task served to obtain post-test measurements of the participants' production of the target sounds. By comparing these measurements with the pre-test measurements, I ask if any changes were made and sustained into the post-test.

5.3.3 Post-processing

Recordings were manually checked for errors or disfluencies and subsequently forced aligned using FAVE (Rosenfelder et al., 2011). FAVE determines phone-level boundaries using the HTK toolkit (Young, 1994) and the CMU American English Pronouncing Dictionary (Carnegie Mellon University, 2008) to determine phonemic representations of words. FAVE was originally fit on American English using the corpus of oral arguments before the Supreme Court of the United States. In some cases, phonemic representations of words are not included in the default dictionary and must be manually added, as was the case for *shit, spritz, sprue,* and *strudel*.

Following forced alignment, phone-level boundaries for the target sibilants and intervening stops were manually corrected using visual inspection of the spectrogram. The sibilant was identified as the interval between the onset and offset of aperiodic frication as evident in the spectrogram. The intervening stop, i.e. following /s/ and preceding /r/ or a vowel, was identified as the interval following the cessation of aperiodic frication as evident in the spectrogram and preceding periodic voicing as evident in the waveform.

5.3.4 Measurements, analyses, & hypotheses

Sibilant measurements

Following manual correction of the sibilant boundaries previously identified by the forced aligner, time-averaged centroid frequency values were extracted for all sibilants using a modified version of a Praat script originally created by DiCanio (2013). Time-averaged values, while removing any dynamic information about the trajectory of centroid frequency production, are more reliable and minimize potential errors in the spectral measurements (Shadle, 2012). Recordings were resampled at 44100 Hz, and time-averaged centroid frequency measurements were calculated using six 15 ms windows with preemphasis at 80 Hz and an examined frequency range from 500 to 12000 Hz. Only the middle 80% of the sibilant was examined, excluding transitions in order to ensure that any potential shifts participants exhibit targeted the sibilant in general and not just the final portion most susceptible to coarticulatory influences. In addition to extracting centroid frequency, the script also extracted time-averaged measurements of the other three spectral moments (standard deviation, skewness, and kurtosis) as well as intensity, interval duration, word duration, and the identity of preceding and following intervals.

Following centroid frequency extraction for all files, the average centroid frequency was then calculated separately for each speaker in each of the different onset environments. Each speaker's RETRACTION RATIO (RR) was then calculated separately for the pre- and post-exposure conditions for each item. The retraction ratio, explained in more detail in Section 2.1, is provided again in Formula 5.1.

$$\text{Retraction Ratio} = \frac{\text{CF of segment} - \text{speaker mean CF of /s/}}{\text{speaker mean CF of /ʃ/} - \text{speaker mean CF of /s/}} \tag{5.1}$$

Recall that a value of 0 suggests that the observed sibilant's centroid frequency value is identical to prevocalic /s/, while a value of 1 suggests that the observed sibilant's centroid

frequency value is identical to prevocalic /ʃ/; thus, a lower value suggests less retraction, while a higher value suggests more retraction.

Two additional measurements were calculated subsequently from the retraction ratio. The first is the DIFFERENCE IN DISTANCE (DID, Babel, 2009), which compares the difference from a baseline measurement to a model talker to the difference from a post-test measurement to a model talker. A positive value indicates that the speaker exhibited convergence and became phonetically more similar to the model talker on the relevant cue during the post-test, while a negative value indicates divergence and demonstrates that speaker became less similar to the model talker. The absolute value of the difference in distance measurements indicates the magnitude of the shift; 0 indicates no change while 1 and -1 exhibit indicate similar shifts toward or away from the model talker, respectively. The formula for difference in distance as applied in this dissertation, examining difference in retraction ratio, is provided in Formula 5.2.

$$\text{Difference in Distance} = |\text{Pre-test RR} - \text{Model RR}| - |\text{Post-test RR} - \text{Model RR}| \quad (5.2)$$

However, while the difference in distance measurement accurately captures many shifts, it can give a misleading indication of divergence in certain scenarios. Specifically, if the speaker exhibits shifts toward the model talker but overshoots the model talker by a significant margin, the difference in distance may return a negative value, suggesting divergence. This is because the difference between the post-test and the model talker may be greater than the difference between the pre-test and the model talker, despite that fact that the shift was in the direction of the model talker's value. Thus, what is acoustically divergence on a technical level may be better described as convergence from both a phonological and social standpoint. To account for this possibility, I introduce a third measurement: the DIRECTION OF SHIFT. As the name suggests, this measurement indicates whether the speaker shifts toward or away from the model talker relative to their baseline measurement. Direction of shift is treated as

binary, coded as 1 (toward the model talker) or 0 (away from the model talker). The formula for direction of shift, examining shifts in retraction ratio between the pre- and post-exposure blocks, is provided in Formula 5.3.

$$\text{Direction of Shift} = \begin{cases} 1 \quad \text{if} \quad (\text{Pre-test RR} > \text{Model RR} \quad \& \quad \text{Post-test RR} < \text{Pre-test RR}) \\ \quad\quad \vee \quad (\text{Pre-test RR} < \text{Model RR} \quad \& \quad \text{Post-test RR} > \text{Pre-test RR}) \\ 0 \quad \text{if} \quad (\text{Pre-test RR} > \text{Model RR} \quad \& \quad \text{Post-test RR} > \text{Pre-test RR}) \\ \quad\quad \vee \quad (\text{Pre-test RR} < \text{Model RR} \quad \& \quad \text{Post-test RR} < \text{Pre-test RR}) \end{cases} \tag{5.3}$$

Analysis

While the centroid frequency for all instances of word-initial /s/ and /ʃ/ was extracted and used to calculate the retraction ratio, and thus the subsequent measurements, difference in distance and direction of shift, only /sCr/ onsets were included for analysis in the present experiment. Each of the three measurements described in the previous section (Retraction Ratio, Difference in Distance, and Direction of Shift) were analyzed separately. First, Retraction Ratio was analyzed to determine whether individuals change their produced centroid frequency of /sCr/ clusters relative to their prevocalic /s/ and /ʃ/ after exposure to a model talker. Secondly, Difference in Distance is analyzed to examine whether participants become more acoustically similar to the model talker after exposure. Thirdly, Direction of Shift was analyzed to ask whether any potential shifts that individuals observe are in the direction of the model talker relative to their baseline.

Separate mixed-effects models were fit on each of the three measurements; for the continuous responses, Retraction Ratio and Difference in Distance, linear mixed-effects models were fit using the `lmer()` function. For Direction of Shift, which is categorical, a generalized linear mixed-effects model with a logit link function was fit using the `glmer()`

function, both from the `lme4` package (Bates et al., 2015) in `R` (R Core Team, 2015). All models included PLACE of articulation (alveolar, velar, and bilabial; Helmert-coded with alveolar as base), RETRACTIONCONDITION (decreased, increased, and hyper-increased, treatment-coded with decreased as base; counterbalanced between subject), and BASELINERELATION (above and below, sum-coded) as fixed effects. Additionally, the model for RETRACTION RATIO included BLOCK (pre- and post-exposure, sum-coded), which was not relevant to the DIFFERENCE IN DISTANCE and DIRECTION OF SHIFT models, as these measurements are calculated by comparing the pre- and post-exposure retraction ratios. Self-reported responses for the demographic categories GENDER (male, female; sum-coded), SEXUALITY (straight, queer; sum-coded), and REGION (midwest, northwest, south, and west, treatment-coded with midwest as base) were included to capture potential social variation. Additionally, each of the social and cognitive scores that are predicted to influence likelihood of convergence were included in each model including MRAS[1], ANXIETY, EMPATHY, and OPENNESS. Each score was scaled, and in the case of ANXIETY inverted, such that a higher value corresponds with an increased likelihood of exhibiting phonetic convergence.

For each measurement, a preliminary model was designed with all two- and three-way interactions between the fixed effects predictors. All interactions that did not reach a significance threshold of 0.05 were pruned from the final model. Additionally, each preliminary model included a maximally specified random effects structure, with by-subject random slopes and intercepts, which were progressively simplified until convergence was achieved. Final models for each measurement are reported in the following section.

1. Separate models were fit on both the composite MRAS score and the toughness subscale, with the composite models outperforming models fit on the toughness subscale. This suggests that endorsements other stereotypes of masculinity besides toughness influence the likelihood of individuals converging toward their interlocutor.

Hypotheses

Building off the findings of Experiment I: Cue Integration (Chapter 3) and Experiment II: Categorization (Chapter 4), which demonstrate that the coarticulatory cues of /s/-retraction are available and useful to the listener in speech perception, the present experiment asks if listener-turned-speakers make adjustments in their speech as a result of exposure to those cues. That is, do they converge toward the spectral cues of retraction? Specifically, the following hypotheses are to be tested:

Hypothesis 1 The first hypothesis proposes that listener-turned-speakers make adjustments in their production of /s/-retraction as a result of the retraction produced by their interlocutor, and do so regardless of the relative degree of retraction that their interlocutor produces. That is, speakers converge toward interlocutors who produce /sCr/ clusters with less retraction than they do and interlocutors who produce /sCr/ clusters with more retraction than they do. Under this hypothesis, participants exhibit significant differences between their pre- and post-exposure retraction ratios for /sCr/ clusters, which result in positive differences in distance and direction of shift measurements. Each of these potential findings would indicate that the speaker converged toward their interlocutor, in this case a model talker, over the course of the task.

Hypothesis 2 The second hypothesis proposes that listener-turned-speakers make adjustments in their production of /s/-retraction as a result of the retraction produced by their interlocutor, but that their likelihood of doing so is constrained by the phonological contrast between /s/ and /ʃ/. A confirmation of this hypothesis would corroborate Nielsen (2011)'s findings for VOT convergence, where speakers only converged toward decreased but never increased VOT. Specifically for this experiment, speakers are predicted to converge toward interlocutors who produce less /s/-retraction than them, as decreasing their relative degree of retraction enhances the contrast between /s/–/ʃ/. In turn, speakers are predicted to either

make no shifts or potentially diverge from interlocutors who produce more /s/-retraction than them, as increasing their relative degree of retraction diminishes the contrast between /s/–/ʃ/. Crucially, this approach first assumes that the sibilant in /sCr/ is phonologically /s/. Secondly, this approach assumes that the maintenance of a phonological contrast influences production even in environments where that contrast is not phonotactically realized. To test this, we look to an effect of BaselineRelation. Individuals with a pre-test value above, i.e. more retracted than, the model talker would support this hypothesis by exhibiting differences between their pre- and post-exposure retraction ratios, and thus positive difference in distance and direction of shift measurements. No such significant shifts, and possibly negative difference in distance and direction of shift measurements, are expected for individuals with a pre-test value below, i.e. less retracted than, the model talker.

Hypothesis 3a In direct contrast to Hypothesis 2, Hypothesis 3a proposes that listener-turned-speakers' likelihood of converging toward the /s/-retraction produced by their interlocutor is constrained by the coarticulatory naturalness of the shift and whether that shifts aligns with an ongoing sound change in progress. A confirmation of this hypothesis would corroborate Zellou et al. (2016)'s findings for coarticulatory nasalization convergence, where speakers only converged toward increased but not decreased nasality, and Pinget (2015)'s findings for fricative devoicing in Dutch, where convergence aligned with an ongoing sound change. Specifically, for this experiment, speakers are predicted to converge toward interlocutors who produce more /s/-retraction than them, as increasing one's relative degree of retraction is more coarticulatorily natural and aligns with /s/-retraction as a change in progress. In contrast, speakers are predicted to either make no shifts or potentially diverge from interlocutors who produce less /s/-retraction than them, as decreasing their relative degree of retraction potentially increases the articulatory effort and contrasts with a sound change in progress. Again, like Hypothesis 2, we look to an effect of BaselineRelation. Evidence for convergence from individuals with a pre-test value below, i.e. less retracted

than, the model talker would support this hypothesis. Evidence of divergence, or null findings of convergence, from individuals with a pre-test value above, i.e. more retracted than, the model talker would support this hypothesis.

Hypothesis 3b Like Hypothesis 3a, Hypothesis 3b proposes that listener-turned-speakers are influenced by coarticulatory properties and their experiences with changes in progress. As a result, participants' likelihood of exhibiting convergence will vary between the different /sCr/ clusters. Specifically, Hypothesis 3b proposes that listener-turned-speakers exhibit more convergence in /str/ clusters, when the model talker is more retracted than they are, than in /skr/ clusters, and more in /skr/ clusters than in /spr/ clusters. To test this, we look to an effect of Place of articulation (alveolar, velar, and bilabial) in conditioning the shifts in retraction ratios or the difference in distance and direction of shift measurements. A confirmation of this hypothesis would provide short-term evidence for a convergence path to sound change, by which sound change is propagated as a result of the accumulation and persistence of short-term shifts. Note that there is not a Hypothesis 2b, whereby place of articulation is predicted to condition convergence toward decreased degrees of retraction, as the same phonological contrast exists (or perhaps doesn't exist) equally for all three clusters.

5.4 Results

The results of this experiment are presented in four parts. First, in Section 5.4.1, I present the results of a model fitted on Retraction Ratio to ask if participants change their produced degree of retraction as a result of exposure to a model talker. In Section 5.4.2, I present the results of a model fitted on Difference in Distance to ask if individuals become more or less similar to the model talker after exposure. In Section 5.4.3, I present the results of a model fitted on Direction of Shift to ask if individuals shifted in the direction of the model talker relative to their baseline. Finally, in Section 5.4.4, I focus on the

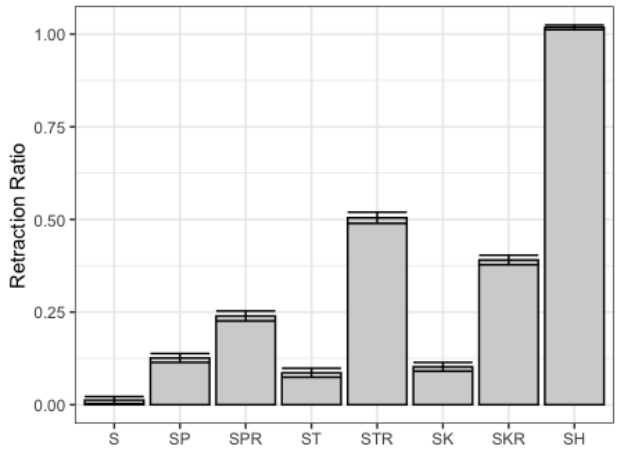


Figure 5.2: Mean and 95% confidence intervals of pre-test retraction ratios for each onset sibilant environment.

individual response patterns exhibited by participants, highlighting the individual variation that may be obfuscated by the community results.

5.4.1 Retraction Ratio results

Before jumping into the results of the model, I first present participants' baseline retraction ratios to demonstrate the patterns individuals come into the lab producing prior to any exposure to the model talker. In Figure 5.2, the mean retraction ratio for each onset sibilant is presented, including prevocalic /s/, /ʃ/, and /sC/ clusters, which are not included in the analysis presented later in this section. Remember that retraction ratio is defined in reference to an individual's mean prevocalic /s/ and /ʃ/, and thus, as expected, prevocalic /s/ has an observed retraction ratio of 0 and prevocalic /ʃ/ has an observed retraction ratio of 1.

As Figure 5.2 demonstrates, participants in the present study from around the continental U.S. generally produce /str/ clusters with a noticeable degree of retraction, such that the mean predicted retraction ratio is 0.50, halfway between prevocalic /s/ and /ʃ/. Furthermore, while the observed retraction ratio is significantly higher than the other onset clusters, /spr/ and /skr/ cluster are produced with notable retraction, with mean values of 0.24 and 0.39,

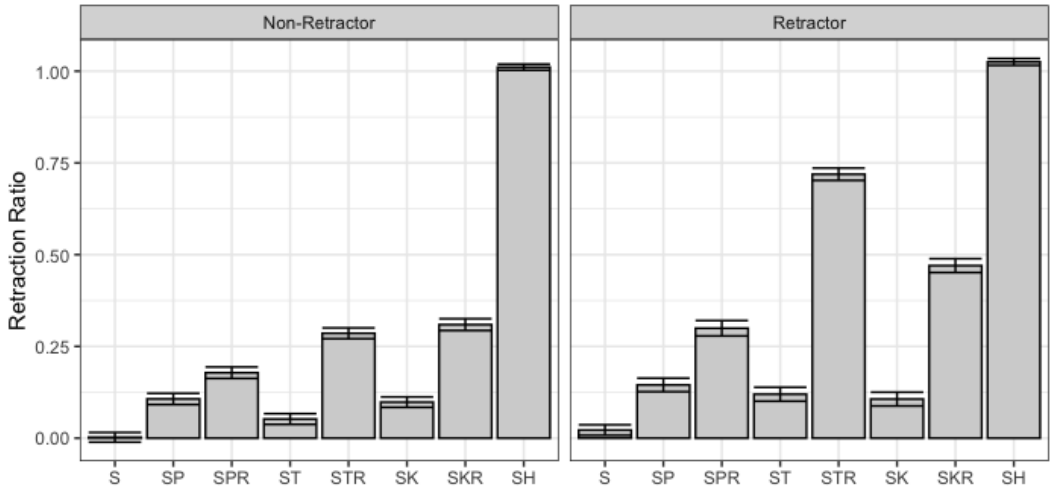


Figure 5.3: Mean and 95% confidence intervals of pre-test retraction ratio for each onset sibilant environment with participants binned as either retractors (RR of /str/ $>$ 0.5) or non-retractors (RR of /str/ $\leq$ 0.5).

respectively. In each place of articulation, it is clear that the /sCr/ cluster is produced with significantly more retraction than its respective /sC/ cluster. In Figure 5.3, retraction ratios for the different onset sibilants are presented separately for individuals whose mean /str/ is above ('retractors') or below ('non-retractors') the group mean of 0.5. In both groups, /sCr/ clusters are significantly more retracted than /sC/ clusters, but for non-retractors in the left-hand panel, there is a less prominent difference between the places of articulation. Specifically, for individuals whose mean /str/ retraction ratio is below 0.5, /skr/ clusters are produced with a degree of retraction similar to /str/ clusters. Furthermore, for the retractors in the right-hand panel, as /str/ increases, so too does /skr/, and to a lesser extent /spr/, yet the /sC/ clusters appear to remain relatively unchanged. These observations provide clear, empirical evidence that /spr/ and /skr/ clusters are sites of retraction, albeit to a lesser degree than /str/ clusters, and should not be disregarded in examinations of /s/-retraction.

Recall that in the present experiment, participants are assigned to one of the retraction conditions, in which the retraction ratio of the model talker is manipulated to be decreased, increased, or hyper-increased relative to the model talker's own baseline. While the con-

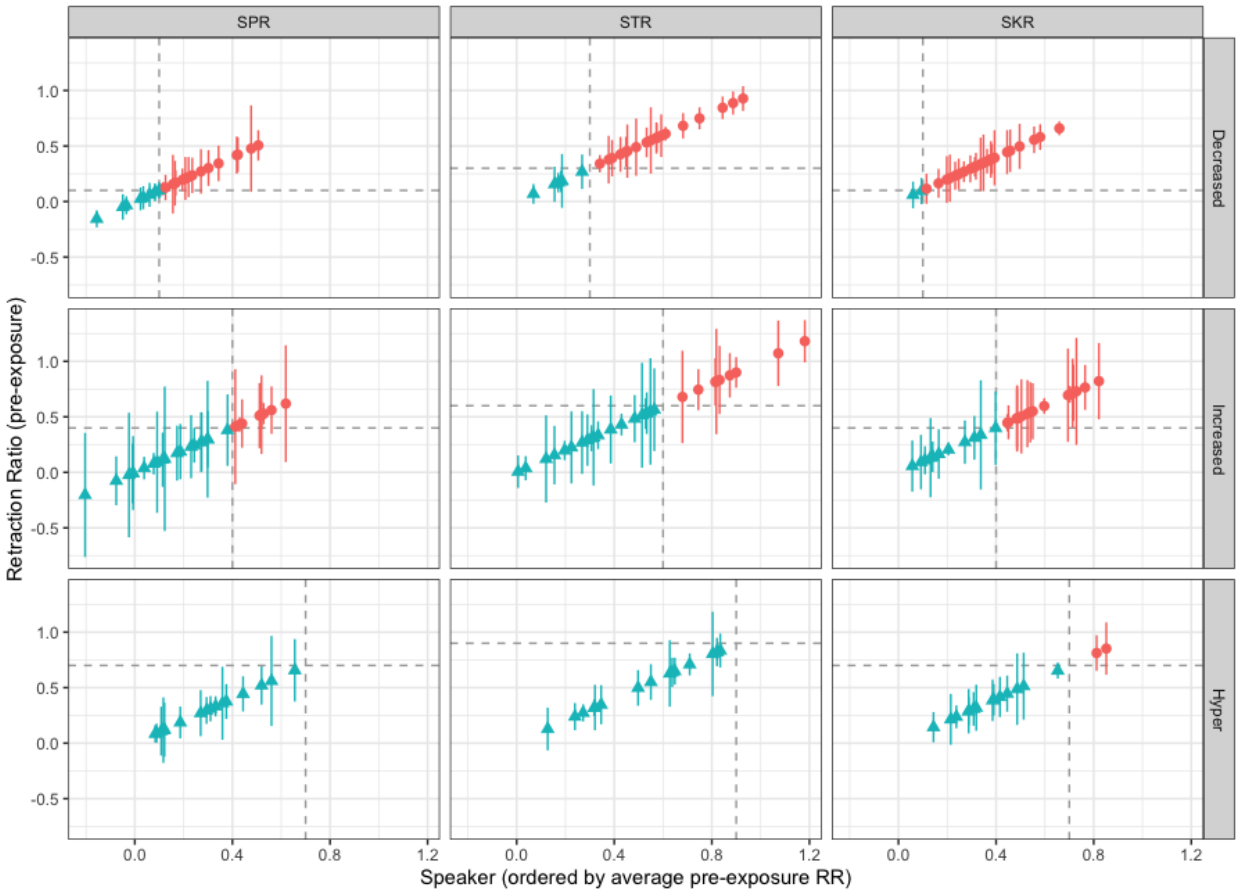


Figure 5.4: Mean and 95% confidence intervals of pre-test retraction ratio for each speaker (x-axis) in the three target clusters (panel column). Participants are presented in ascending order, separated by retraction condition (panel row). The model talker's value is indicated by the dashed gray lines intersecting both the x- and y-axes. Participants whose baseline is more retracted than the model talker for the relevant cluster are indicated in red; participants whose baseline is less retracted than the model talker are indicated in teal.

ditions were designed based off the model talker's own production values, they were also informed by the participants of previous experiments in the Phonology Laboratory at the University of Chicago. Thus, the expectation was that the majority of participants would exhibit a baseline above the model talker in the decreased retraction condition but below the model talker in the increased retraction condition. In Figure 5.4, each participant's mean baseline retraction ratio is presented, separated by condition and /sCr/ cluster. Individuals whose retraction ratio is above the model talker for a given cluster/retraction condition are

indicated in red, while those below the model talker are indicated in teal. As illustrated in Figure 5.4, most participants in the decreased retraction condition have a baseline above the model talker, while more than half of the participants in the increased retraction condition have a baseline below the model talker. However, the number of individuals above the model talker in the increased retraction condition in /str/ and /skr/ clusters motivated the post-hoc addition of the hyper-increased retraction condition, in which, as illustrated in the bottom row of Figure 5.4, nearly every participant exhibited a retraction ratio below the model talker. Furthermore, while the expected trends for an individual's baseline relative to the model talker hold in each condition, as shown in Figure 5.4, the variation, especially in the increased retraction condition, motivated the inclusion of BaselineRelation. BaselineRelation simply defines the position of the participant's baseline retraction ratio relative to the model talker, coded as 'above' or 'below'. Thus, we can account for potential differences within the same condition. For example, if two individuals in the increased retraction condition both converge toward the model talker, then the participant with a baseline above would decrease their relative degree of retraction, while the participant below would increase their relative degree of retraction.

With an informed perspective of the baseline patterns participants bring into the laboratory, I turn to an analysis of retraction ratios in both the pre- and post-exposure blocks to ask if individuals shift in response to the model talker. The final model for retraction ratio selected after progressive simplification of the random effects structure in order to achieve model convergence is presented in `lme4` format in Formula 5.4.

$$\begin{aligned}&\textsc{RetractionRatio} \sim (\textsc{Cluster}+\textsc{RetractionCondition}+\textsc{BaselineRelation}+\textsc{Block})^3+\\&\textsc{Block} * (\textsc{Region}+\textsc{Gender}+\textsc{Sexuality}+\textsc{MRAS}+\textsc{Empathy}+\textsc{Openness}+\textsc{Anxiety})\\&(1+\textsc{Cluster}|\textsc{Subject})+(1|\textsc{Word})\end{aligned} \tag{5.4}$$

Table 5.3: Model predictions for all significant main effects and interactions in retraction ratio as a result of exposure to a model talker with varied degrees of retraction, N=2854. Cluster1 indicates the first contrast for Cluster, i.e. /str/ vs the combined /spr/ and /skr/, and Cluster2 indicates the second contrast, i.e. /skr/ vs. /spr/. A positive value indicates a higher, i.e. more /ʃ/ like, retraction ratio. Complete model predictions including variables and interactions that did not reach a significance threshold of 0.05 are included in the Appendix as Table 5.3.

	Est.	*SE*	*t*	*p*
(Intercept)	0.09	0.04	1.82	0.069
Cluster1	0.19	0.05	3.82	< 0.001
Increased	0.17	0.05	3.03	0.002
Hyper	0.30	0.06	5.45	< 0.001
Block	0.19	0.04	5.21	< 0.001
BaselineRelation	0.32	0.03	9.75	< 0.001
Increased:Block	−0.17	0.04	−3.97	< 0.001
Hyper:Block	−0.18	0.04	−4.52	< 0.001
Block:BaselineRelation	−0.17	0.04	−4.34	< 0.001
Cluster1:Block:BaselineRelation	−0.12	0.05	−2.23	0.025
Cluster2:Block:BaselineRelation	−0.13	0.07	−2.04	0.041
Block:Queer	−0.08	0.02	−3.68	< 0.001
Block:MRAS	−0.04	0.01	−3.40	< 0.001
Block:Anxiety	0.04	0.01	4.18	< 0.001

This model asks how individuals may change their centroid frequency of an /sCr/ relative to their mean prevocalic /s/ and /ʃ/ as a result of Cluster identity (the different places of articulation), Block (before or after exposure to the model talker), RetractionCondition (the degree of /s/-retraction exhibited by the model talker), and BaselineRelation (the relationship of the participant's baseline relative to the model talker). The social and cognitive predictors are included to account for any individual variability, especially given the between-subject design of the present experiment. The significant effects and interactions of the retraction ratio model are presented in Table 5.3.

Figure 5.5 plots the retraction ratio by Cluster (/spr/, /str/, and /skr/), block (pre- or post-exposure), and RetractionCondition (decreased, increased, or hyper-increased). Visual inspection of the figure first illustrates what was also suggested in Figure 5.3, that

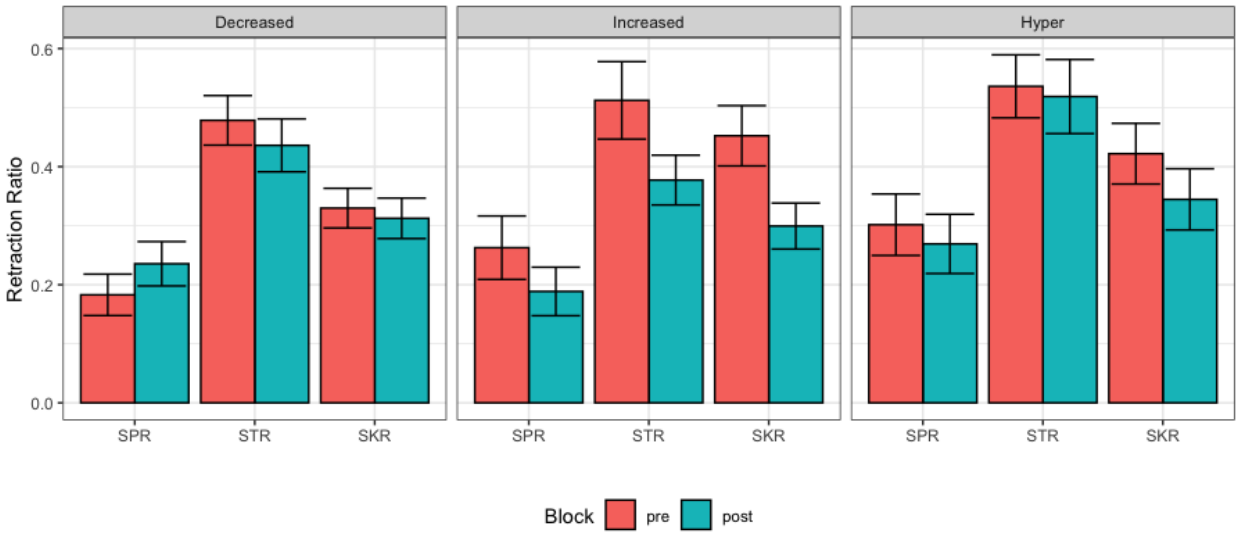


Figure 5.5: Mean and 95% confidence intervals of the retraction ratio of each target clusters (x-axis) in the pre- and post-test (color: pre-exposure = red, post-exposure = teal). The three retraction conditions (decreased, increased, and hyper-increased) are presented separately by column panel.

/str/ clusters are produced with the most retraction, followed by /skr/ clusters, followed by /spr/ clusters. This is confirmed by the model, with a significant main effect of Cluster1. Remember that as Cluster is Helmert-coded, the first comparison is between /str/ and the mean of /spr/ and /skr/. Thus, a positive effect Cluster1 demonstrates that, all else being equal, /str/ clusters are produced with a significantly higher degree of retraction, i.e. more /ʃ/-like, than /spr/ and /skr/ clusters ($t = 1.82, p < 0.001$). While visual inspection of Figure 5.5 as well as the previously examined Figure 5.3 suggest that /skr/ is produced with a higher retraction ratio than /spr/, this was not supported by the model ($t = -0.03, p = 0.970$).

As demonstrated in Figure 5.5, there is an overall tendency that the retraction ratio values are lower, i.e. less /ʃ/-like, in post-test than the pre-test. As a general trend, this holds across different RetractionConditions and Places of articulation, with the exception of /spr/ clusters in the decreased RetractionCondition, and is especially robust in the increased and hyper-increased RetractionCondition for all clusters. This observation is confirmed in the model, not as an effect of Block, in which the post-exposure block

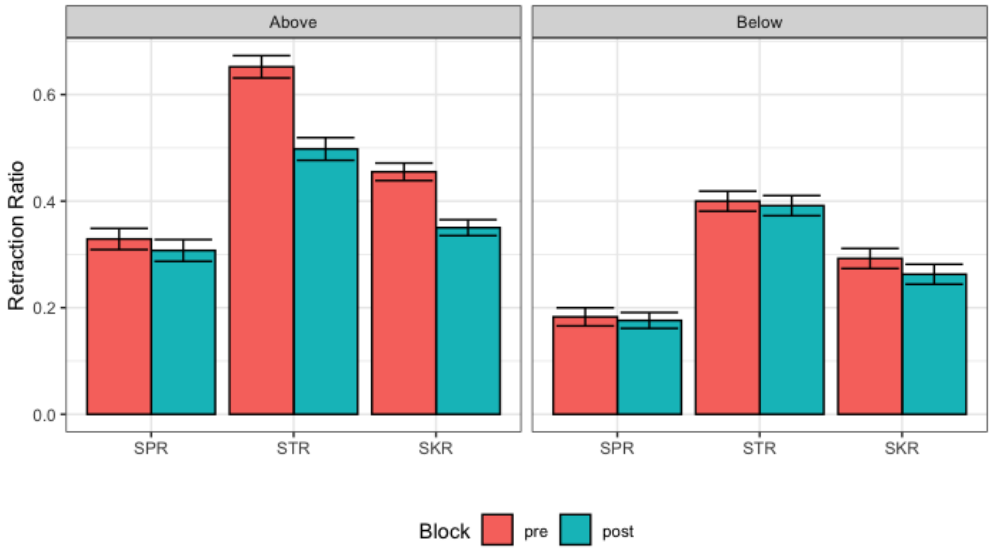


Figure 5.6: Mean and 95% confidence intervals of the retraction ratio of each target clusters (x-axis) in the pre- and post-test (color: pre-exposure = red, post-exposure = teal). The three retraction conditions (decreased, increased, and hyper-increased) are presented separately by column panel. Retraction ratios are presented separately for individuals whose baseline was greater (i.e. more retracted) than the model talker for a given item (left panel) and those whose baseline was less (i.e. less retracted) than the model talker (right panel)

is predicted to be characterized by a higher, i.e. more /ʃ/-like, centroid frequency ($t = 5.21, p < 0.001$), but rather in the interaction of Block and increased and hyper-increased RetractionCondition. For both retraction conditions, the model predicts a decreased retraction ratio in the post-exposure block relative to the pre-exposure block (Increased: $t = -3.97, p < 0.001$; Hyper: $t = -4.52, p < 0.001$).

Figure 5.6 adds in the remaining primary predictor of the model fit on on retraction ratio, swapping RetractionCondition out in favor of BaselineRelation. Visual inspection of the figure first and foremost illustrates that, all else being equal, individuals with a baseline retraction ratio above the model talker are observed to have a retraction ratio higher, i.e. more/ʃ/-like, than those with a baseline retraction ratio below the model talker, which is true by definition. Unsurprisingly, this is supported by the main effect of BaselineRelation ($t = 9.75, p < 0.001$). Secondly, visual inspection of Figure 5.6 suggests that larger differences are observed between the pre- and post-exposure retraction ratios for /str/ and /skr/ clusters

for individuals whose baseline is above the model talker, while less discernible shifts are observed for individuals below the model talker. This is confirmed by the model with a significant two-way interaction of BaselineRelation and Block ($t = -4.32, p < 0.001$), as well the significant three-way interaction of Cluster, BaselineRelation, and Block (Cluster1: $t = -2.23, p = 0.025$; Cluster2: $t = -2.04, p = 0.041$). These findings suggest that individuals are significantly more likely to decrease their retraction ratio, especially in /str/ and /skr/ clusters, as a result of their exposure to a model talker with a retraction ratio less than their own. Such shifts are thus toward the model talker, providing evidence for phonetic convergence in these environments.

Additionally, the model suggests that while none of the social or cognitive indicators are significant predictors of an individual's retraction ratio production on their own, individuals who identify as queer (or non-straight, including gay, lesbian, bisexual, asexual, queer, or questioning) and individuals who more strongly endorse masculine stereotypes, perhaps an unlikely pairing, are more likely to exhibit a decreased retraction ratio after exposure to the model talker (Block:Sexuality: $t = -3.68, p < 0.001$; Block:MRAS: $t = -3.40, p < 0.001$). These interactions are not explicitly measures of convergence, as they do not include any indication of whether their respective participants had a baseline retraction ratio above the model talker such that the decreased value observed in the post-exposure block would exhibit a shift toward the model talker. Nonetheless, they align with the more robust pattern of convergence observed. In contrast, the model suggests that individuals who are less anxious are more likely to produce an increased retraction ratio following exposure to the model talker ($t = 4.18, p < 0.001$), suggesting a potential instance of divergence.

5.4.2 Difference in Distance results

The retraction ratio results in the previous section illustrate that participants' degree of retraction was influenced by exposure to a model talker with manipulated degrees of retraction. The effect of BaselineRelation in particular illustrates that the nature of that influence is

dependent on the participant's own baseline degree of retraction relative to the model talker. In the present section, I turn to a second model fit on DIFFERENCE IN DISTANCE, which asks not what participant's retraction ratio patterns produced were, but rather whether individuals became more or less similar to the model talker in retraction ratio in the post-exposure block. The final model for difference in distance selected after progressive simplification of the random effects structure in order to achieve model convergence is presented in `lme4` format in Formula 5.5.

$$\begin{aligned}&\text{DifferenceInDistance} \sim (\text{Cluster}+\text{RetractionCondition}+\text{BaselineRelation})^3+\\&(\text{Region}+\text{Gender}+\text{Sexuality}+\text{MRAS}+\text{Empathy}+\text{Openness}+\text{Anxiety})+\\&(1+\text{Cluster}|\text{Subject})+(1|\text{Word})\end{aligned} \tag{5.5}$$

This model asks how participants become more similar to the model talker in retraction ratio for /sCr/ clusters in their post-exposure production relative to their pre-exposure values. Like the model fit on retraction ratio, the difference in distance model includes CLUSTER identity, i.e. the different places of articulation, RETRACTIONCONDITION, i.e. the degree of /s/-retraction exhibited by the model talker, and BASELINERELATION, i.e. the relationship of the participant's baseline relative to the model talker. Unlike the model fit on retraction ratio, the model for difference in distance does not include BLOCK as a predictor, as difference in distance by definition compares the pre- and post-exposure measurements. Once again, the social and cognitive predictors are included to account for any individual variability, especially given the between-subject design of the present experiment. The significant effects and interactions of the difference in distance model are presented in Table 5.4 and the complete results in the Appendix as Table A.5.

Difference in distance captures convergence as it asks whether the distance between the pre-test and the the model talker is greater than the distance between the post-test and the model talker. That is, convergence is said to have occurred when the distance from the

Table 5.4: Model predictions for all significant main effects and interactions in difference in distance as a result of exposure to a model talker with varied degrees of retraction, N=816. Cluster1 indicates the first contrast for Cluster, i.e. /str/ vs the combined /spr/ and /skr/, and Cluster2 indicates the second contrast, i.e. /skr/ vs. /spr/. A positive value indicates greater convergence toward the model talker. Complete model predictions including variables and interactions that did not reach a significance threshold of 0.05 are included in the Appendix as Table A.5.

	Est.	*SE*	*z*	*p*
(Intercept)	−0.07	0.03	−1.97	0.049
BaselineRelation	0.07	0.03	2.08	0.037
Hyper:BaselineRelation	−0.13	0.05	−2.19	0.028
Hyper:Cluster1:BaselineRelation	−0.24	0.12	−2.09	0.036
Northeast	0.05	0.02	1.99	0.045

model talker shrinks, which would yield a positive difference in distance value. Conversely, if the post-test distance is greater than the pre-test, this yields a negative difference in distance value, suggesting divergence. For the present experiment, we analyze the distance in retraction ratios, which unlike previous examinations of difference in distance (e.g. Babel, 2012; Pardo et al., 2013; Zellou et al., 2016), is not a purely acoustic measurement, and thus does not capture acoustic convergence, but rather is informed by an individual's phonological contrast between /s/ and /ʃ/ and captures the convergence toward a phonetic/phonological system. This is crucial because /s/-retraction is not the realization of a particular raw frequency goal for /s/ in /str/ clusters, but rather the realization of /s/ as approaching the individual's phonological category for /ʃ/.

Figure 5.7 summarizes the likelihood of convergence, plotting difference in distance by Cluster (/spr/, /str/, and /skr/), RetractionCondition (decreased, increased, or hyper-increased), and BaselineRelation (above or below). Initial inspection of the left-hand panel for the participants with a baseline retraction ratio above the model talker illustrates notable variation in the predicted difference in distance, with positive values predicted in /str/ and /skr/ clusters in the decreased and increased retraction conditions, but negative or zero values predicted in /spr/ clusters and the hyper-increased retraction con-

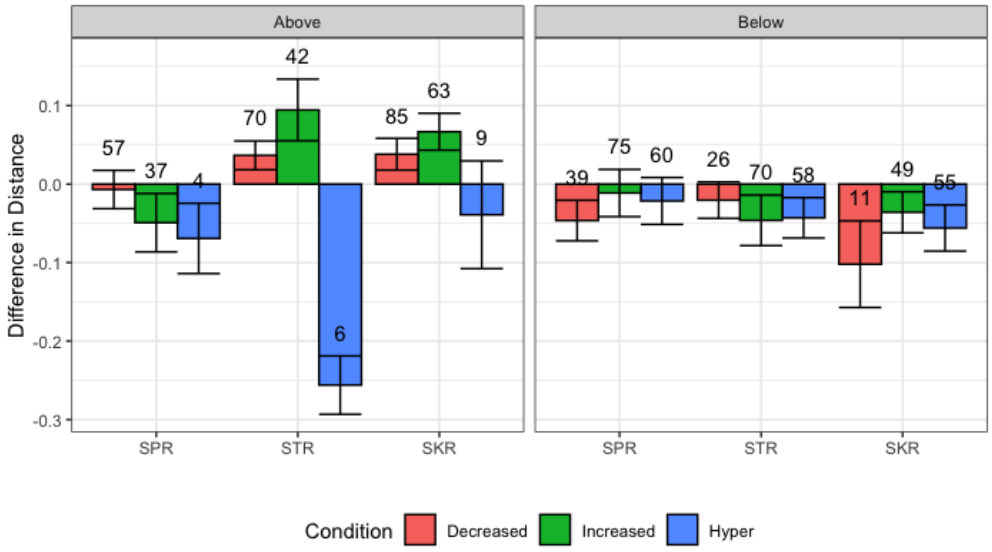


Figure 5.7: Mean and 95% confidence intervals of difference in distance for each target clusters (x-axis) in the three retraction conditions (color: decreased = red, increased = green, hyper-increased = blue). Difference in distance is presented separately for individuals whose baseline retraction ratio is greater than the model talker for a given item (left panel) and those whose retraction ratio is less than the model talker (right panel). As observations did not evenly fall in each condition, counts are noted for each bar.

dition. On the right-hand panel for the participants with a baseline retraction ratio below the model talker, negative values are predicted nearly across the board. These observations suggest that individuals below the model talker diverge, while speakers above the model talker converge, but only in certain conditions. This is supported by the model, with higher difference in distance values predicted for individuals with a mean baseline retraction ratio above the model talker ($t = 2.08, p = 0.037$). Furthermore, the model predicts a subsequent decrease in the predicted difference of distance values for the above BaselineRelation, the (hyper)-increased RetractionCondition, and in /str/ clusters (Hyper:Above: $t = -2.19, p = 0.028$; Cluster1:HyperAbove: $t = -2.09, p = 0.036$). However, it is worth noting observations are not evenly distributed between the different pairings of BaselineRelation and RetractionCondition. Specifically, while there are approximately equal numbers of participants above and below the model talker in the increased RetractionCondition, participants are highly skewed to be above the model talker in

the decreased retraction condition and below the model talker in the hyper-increased retraction condition. For this reason, I have noted the number of observations for each point in Figure 5.7 and these interactions with few observations should be interpreted cautiously. Specifically, the interaction between the above BaselineRelation and the hyper-increased RetractionCondition should interpreted with additional caution.

A post-hoc model designed to account for this skewed distribution was identical to the model presented in Formula 5.5, but RetractionCondition was excluded. Additionally, the contrast coding of Cluster was redesigned as /str/ and /skr/ pattern to the exclusion of /spr/ clusters, and thus /spr/ was first compared to the mean of /str/ and /skr/ and then /skr/ was compared to /str/. From this model, only two predictors emerge as significant. First, the intercept is again negative, suggesting that, all else being equal, individuals are more likely to diverge from the model talker than converge (intercept: $t = -2.11, p = 0.035$). Secondly, individuals above the model talker for /str/ and /skr/ clusters are more likely to converge (Cluster1:BaselineRelations: $t = 2.32, p = 0.020$). While this post-hoc model should be approached cautiously, these predictions capture what appears to be illustrated visually: Individuals who start with a baseline retraction ratio above the model talker are more likely to converge, demonstrated by a positive difference in distance, than individuals who start with a baseline retraction ratio below the model talker. However, as mentioned in Section 5.3.4, it is possible that if a participant overshoots the model talker, difference in distance may register divergence when, from a phonological and social standpoint, the participant may be understood to exhibit convergence.

Finally, turning to the social predictors, only Region emerged as significant in the initial model, with individuals raised in the northeastern United States slightly more likely to decrease their retraction distance from the model talker over the course of the interaction relative to individuals raised in the Midwest ($t = 1.99, p = 0.045$).

5.4.3 Direction of Shift results

The previous two models fit on retraction ratio and difference in distance appear to offer conflicting results: The retraction ratio model suggests that individuals more retracted than the model talker exhibit convergence, while individuals less retracted than the model talker do not appear to shift significantly (Figure 5.6); on the other hand, the difference in distance model suggests that individuals more retracted than the model talker exhibit a variety of patterns, while individuals less retracted exhibited subtle, but consistent divergence (Figure 5.7). The third and final model of this experiment turns to direction of shift, which is a binary measurement that essentially asks if any shift is observed between the pre- and post-test, is it toward the model talker? This is especially necessary considering the potential that if any individual overshoots the target of the model talker, they may end up with a less similar retraction ratio and thus a negative difference in distance value that suggests divergence. The final model for direction of shift, selected after progressive simplification of the random effects structure to achieve model convergence, is presented in `lme4` format in Formula 5.6.

$$\begin{aligned}&\textsc{Direction of Shift} \sim (\textsc{Cluster}+\textsc{RetractionCondition}+\textsc{BaselineRelation})^3+\\&(\textsc{Region}+\textsc{Gender}+\textsc{Sexuality}+\text{MRAS}+\textsc{Empathy}+\textsc{Openness}+\textsc{Anxiety})*\\&(1+\textsc{Cluster}|\textsc{Subject})\end{aligned} \tag{5.6}$$

This model asks if participants shift their post-exposure retraction ratio of /sCr/ clusters toward or away from the model talker relative to their pre-exposure values. Like the model fit on retraction ratio and difference in distance, the direction of shift model includes Cluster identity, RetractionCondition, and BaselineRelation. Like the difference in distance model, Block is not included, as it is relevant to calculating but not characterizing direction of shift. The significant effects and interactions of the retraction ratio model are presented

Table 5.5: Model predictions for all significant mains effects and interactions in direction of shift as a result of exposure to a model talker with varied degrees of retraction, N=816. Cluster1 indicates the first contrast for Cluster, i.e. /str/ vs the combined /spr/ and /skr/, and Cluster2 indicates the second contrast, i.e. /skr/ vs. /spr/. A positive value indicates greater convergence toward the model talker. Complete model predictions including variables and interactions that did not reach a significance threshold of 0.05 are included in the Appendix as Table 5.5.

	Est.	*SE*	*z*	*p*
(Intercept)	0.86	0.41	2.31	0.021
Increased	−1.48	0.47	−3.09	0.002
Hyper	−1.86	0.47	−3.90	< 0.001
BaselineRelation	−0.89	0.41	−2.15	0.031
Increased:Cluster1	1.42	0.75	1.90	0.050
Hyper:Cluster1	1.59	0.73	2.16	0.031
Increased:BaselineRelation	3.34	0.60	5.55	< 0.001
Cluster1:BaselineRelation	1.52	0.74	2.05	0.040
Northeast	0.99	0.31	3.13	0.002
West	0.86	0.38	2.25	0.024

in Table 5.5.

Figure 5.8 summarizes the likelihood of convergence, plotting direction of shift by Cluster (/spr/, /str/, and /skr/) and RetractionCondition (decreased, increased, or hyper-increased). In contrast to Figure 5.7, the majority of observations appear to indicate convergence, with a direction of shift value over 0.5, indicating an increased likelihood of convergence (a direction of shift value below 0.5 indicates that divergence is more likely). This is supported by the model with a positive intercept, suggesting that, all else being equal, participants are more likely to shift toward the model talker than shift away ($z = 2.32, p = 0.021$). This contrasts with the difference in distance findings, suggesting that many of the participants may have overshot the model talker, especially in /spr/ clusters and the decreased retraction condition.

Further visual inspection of Figure 5.8 suggests that convergence is more likely in the decreased and increased retraction conditions than in the hyper-increased retraction condition. In fact, in the hyper-increased retraction condition, trends toward divergence are observed.

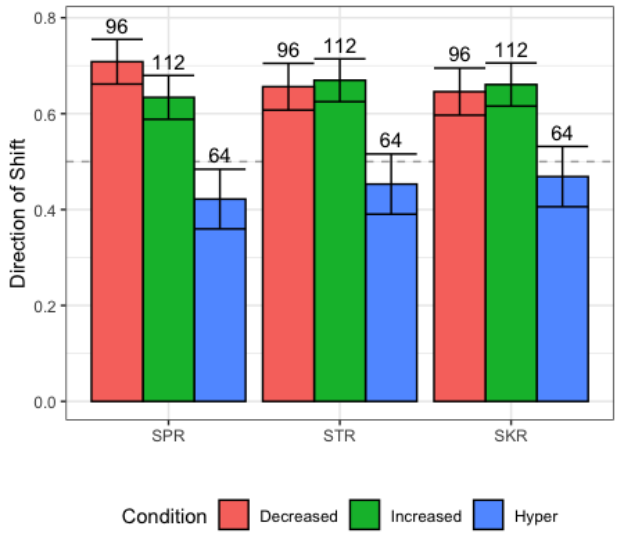


Figure 5.8: Mean and 95% confidence intervals of direction of shift for each target clusters (x-axis) in the three retraction conditions (color: decreased = red, increased = green, hyper-increased = blue). The dashed line at 0.5 separates predicted convergence (DoS > 0.5) from predicted divergence (DoS < 0.5). As observations did not evenly fall in each condition, counts are noted for each bar.

This is supported by the model with less convergence predicted in the hyper-increased retraction condition relative to the decreased retraction condition ($z = -3.90, p < 0.001$), and, perhaps less expectedly from Figure 5.8, a similar prediction of decreased convergence in the increased retraction condition ($z = -3.09, p = 0.002$). However, this effect becomes clearer when considering the complex way in which BaselineRelation influences convergence patterns, particularly in the increased retraction condition where individuals are more evenly split between a baseline retraction ratio above and below the model talker.

Figure 5.9 adds BaselineRelation (above or below) to highlight an interesting pattern: Individuals with a baseline above the model talker are more like to converge in the increased or hyper-increased retraction condition, while individuals with a baseline below the model talker are more likely to converge in the decreased retraction condition. This appears to go beyond the skewed nature of the RetractionCondition-BaselineRelation pairings. That is, in the increased retraction condition with significant numbers above and below the

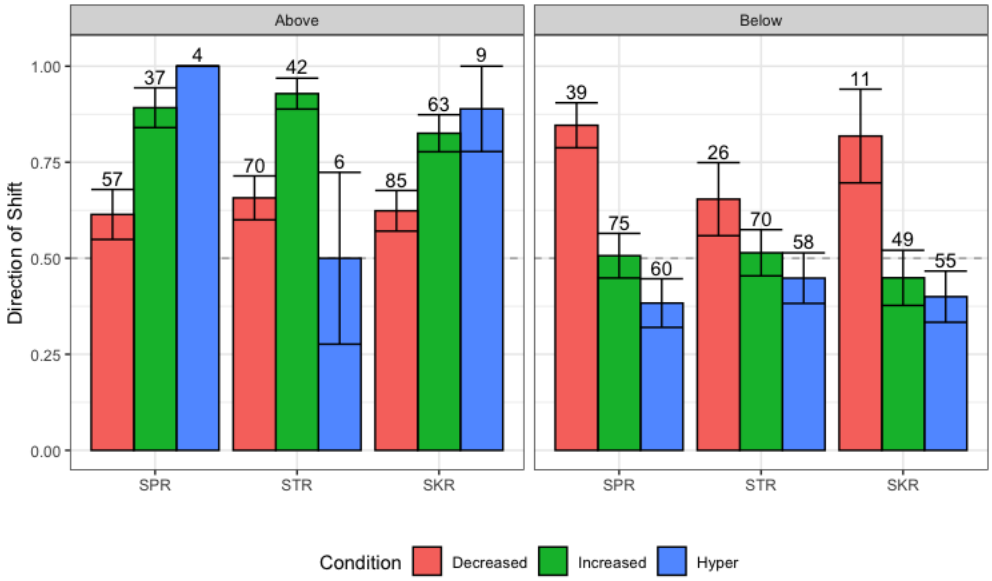


Figure 5.9: Mean and 95% confidence intervals of direction of shift for each target clusters (x-axis) in the three retraction conditions (color: decreased = red, increased = green, hyper-increased = blue). Direction of shift is presented separately for individuals whose baseline retraction ratio is greater than the model talker for a given item (left panel) and those whose retraction ratio is less than the model talker (right panel). As observations did not evenly fall in each condition, counts are noted for each bar. The dashed line at 0.5 separates predicted convergence (DoS $>$ 0.5) from predicted divergence (DoS $<$ 0.5).

model talker, convergence is robustly observed for individuals with a baseline above the model talker. No significant shifts are observed for those below the model talker. In the hyper-increased retraction condition, with the vast majority of observations having a baseline below the model talker, the trend is toward divergence rather than convergence. Meanwhile, in the decreased retraction condition, convergence is observed for individuals with a baseline both above and below the model talker, but the likelihood of convergence increases when their baseline is below the model talker. Thus, the previously discussed main effects of RetractionCondition, predicting less convergence in the increased and hyper-increased retraction conditions relative to the decreased retraction condition, make more sense after observing the more consistent, albeit smaller, shifts in the decreased retraction condition regardless of the the speaker's baseline. In comparison, the increased (and, less reliably

due to the sparse number of observations, the hyper-increased) condition is characterized by stronger prediction of convergence for individuals with a baseline above the model talker, manifested in the model by the interaction of increased retraction condition with above baseline relation ($z = 3.34, p < 0.001$). These findings suggest that while convergence is the norm in the direction, albeit not the size, of the shifts, there appears to be some additional influence on when and where convergence is observed, which is discussed in further detail in Section 5.5.

Finally, like in the difference in distance model, only REGION out of all the social predictors emerged as significant in the direction of shift model, with individuals raised in the northeastern and western United States more likely to shift their retraction ratio in the direction of the model talker (Northeast: $t = 3.13, p = 0.002$; West: $t = 2.25, p = 0.024$).

5.4.4 Individual results

Each of the different models presented earlier in this chapter asks, whether given a similar set of conditions, a community of speakers will be more or less likely to converge toward a model talker's relative degree of /s/-retraction. I've asked if this is influenced by the patterns that the model talker exhibits or the participants' own patterns that they bring into the laboratory. In the final portion of this section, I examine the individuals that collectively make up the subject pool, asking not what patterns emerge from their pooled results, but rather what the individual variation between participants can tell us about convergence toward manipulated degrees of /s/-retraction.

Figure 5.10 plots each individuals pre- and post-test retraction ratios for each of the three target CLUSTERS (/spr/, /str/, and /skr/). Individual results are presented separately for each of the three RETRACTIONCONDITIONS (decreased, increased, and hyper-increased). For each individual, ordered by their baseline retraction ratio, I ask whether there are notable shifts between their pre- and post-exposure retraction ratio values. Furthermore, the

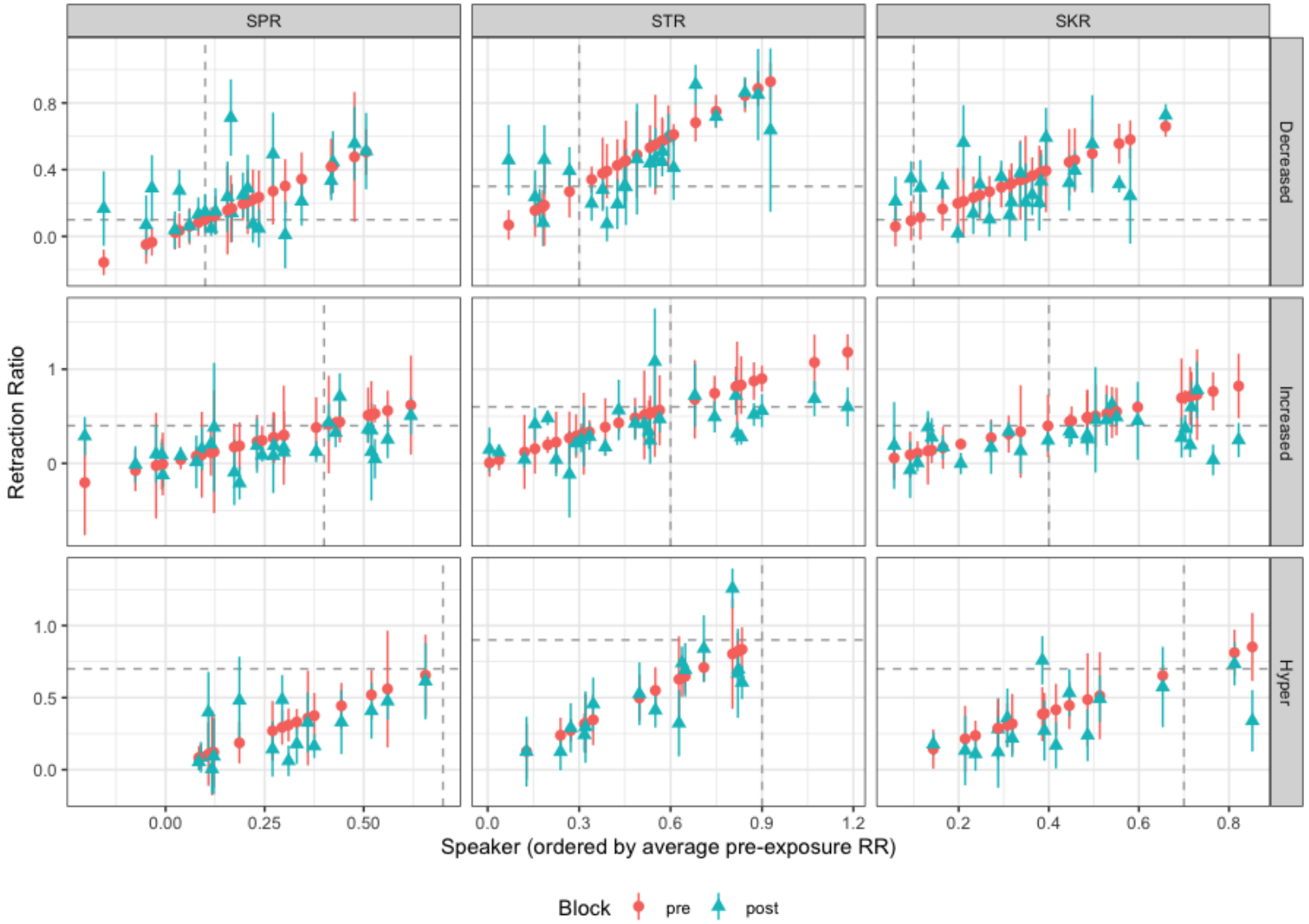


Figure 5.10: Mean and 95% confidence intervals of pre- (red) and post-exposure (teal) retraction ratios for each speaker in the three target clusters (panel column). Participants are presented in ascending order, separated by condition (panel row). Individuals who pre- and post-exposure confidence intervals do not overlap exhibited significant shifts between blocks. The model talker's value is included for reference, indicated by the dashed gray lines intersecting both the x- and y-axes.

model talker's retraction ratio values have been indicated, and, with these values as our guideposts, I ask whether the potential shifts approach and/or overshoot the model talker. If the confidence intervals for the pre- and post-exposure mean do not overlap, this indicates that an individual's mean retraction ratio has shifted significantly (using a 0.05 threshold of significance).

With each individual's pre- and post-exposure values for each cluster, Figure 5.10 can be a lot to take in at first glance. Perhaps most noteworthy is the observation that the majority of confidence intervals overlap, suggesting that although convergence may be predicted for the group mean, most individuals are not making shifts large enough to be significant in their own right, due to the relative low number of observations per individual. Furthermore, while substantial variation can be observed with the direction of the shifts, with post-exposure means frequently observed both above and below the pre-exposure values, there are some noticeable trends. One trend is that individuals with a baseline measurement below the model talker, especially in the decreased retraction condition, are more likely to produce a post-exposure mean above, i.e. more retracted than, their pre-exposure mean. Conversely, individuals with a baseline measurement above the model talker, especially in the increased retraction condition (and though exceedingly rare, in the hyper-increased retraction condition), are more likely to exhibit a post-exposure mean below, i.e. less retracted than, their pre-exposure mean, and more likely for that difference to reach the threshold of significance. This observation mirrors the direction of shift results, with a clearer understanding of the variation of patterns exhibited.

Figure 5.11 focuses on /str/ clusters in increased retraction conditions, essentially zooming in on the centermost panel of Figure 5.10. Participants are ordered in sequence by their baseline retraction ratio, such that they are evenly spaced to ease the examination of each individual's values, and as such, the distances between participants do not represent the differences between their baseline retraction ratios. The main takeaway from Figure 5.11

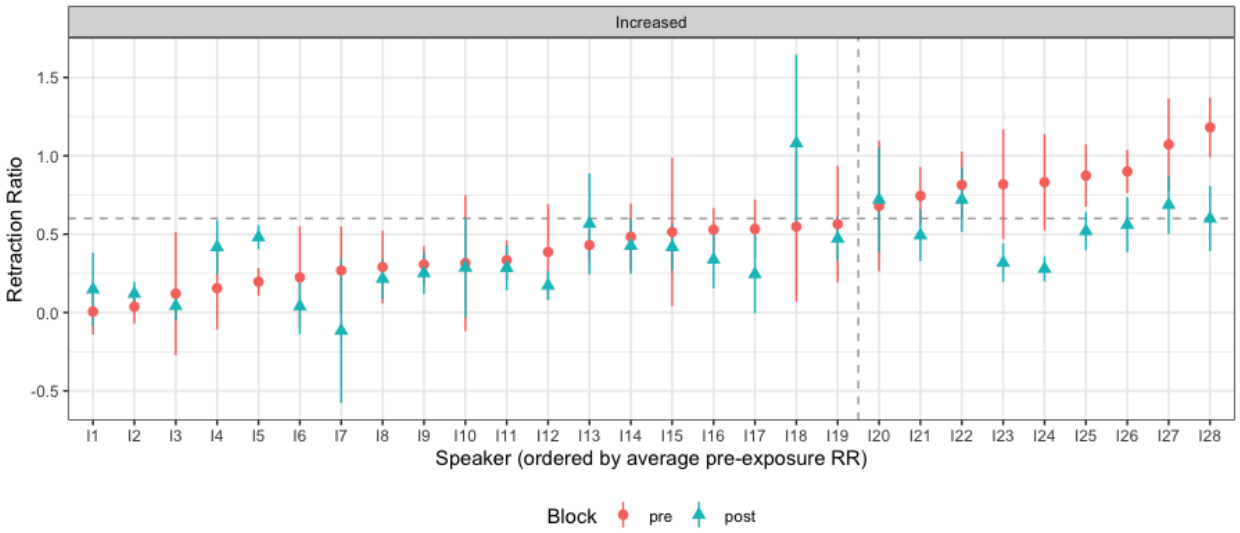


Figure 5.11: Mean and 95% confidence intervals of /str/ pre- (red) and post-exposure (teal) retraction ratios for each speaker. The model talker's value is included for reference, indicated by the dashed gray lines intersecting both the x- and y-axes.

is the near categorical differences in shifts observed as a result of baseline relation. For individuals with a baseline /str/ more retracted than the model talker (plotted to the right of the dashed line) eight out of nine individuals exhibit a shift toward the model talker by lowering their degree of retraction. For five out of nine (I23, I24, I25, I26, and I28), there is no overlap between the confidence intervals for their pre- and post-exposure mean, which suggests that the individuals made a significant shift as a result of the exposure to the model talker. Furthermore, six out of the nine (I21, I23, I24, I25, I26, and I28) overshot the model talker, ending with a post-exposure retraction ratio for /str/ less than 0.6. In contrast, for individuals who began with a baseline less retracted than the model talker (plotted to the left of the dashed line), only six out of nineteen exhibited a shift toward the model talker. Only one individual reached the threshold of significance (I5), and only one individual overshot the model talker (I18).

While some of the observations of individual shift patterns noted for /str/ clusters in the increased retraction hold for clusters and other conditions, convergence patterns for other clusters and conditions are less categorical. Figure 5.12 plots the individual retraction ratio

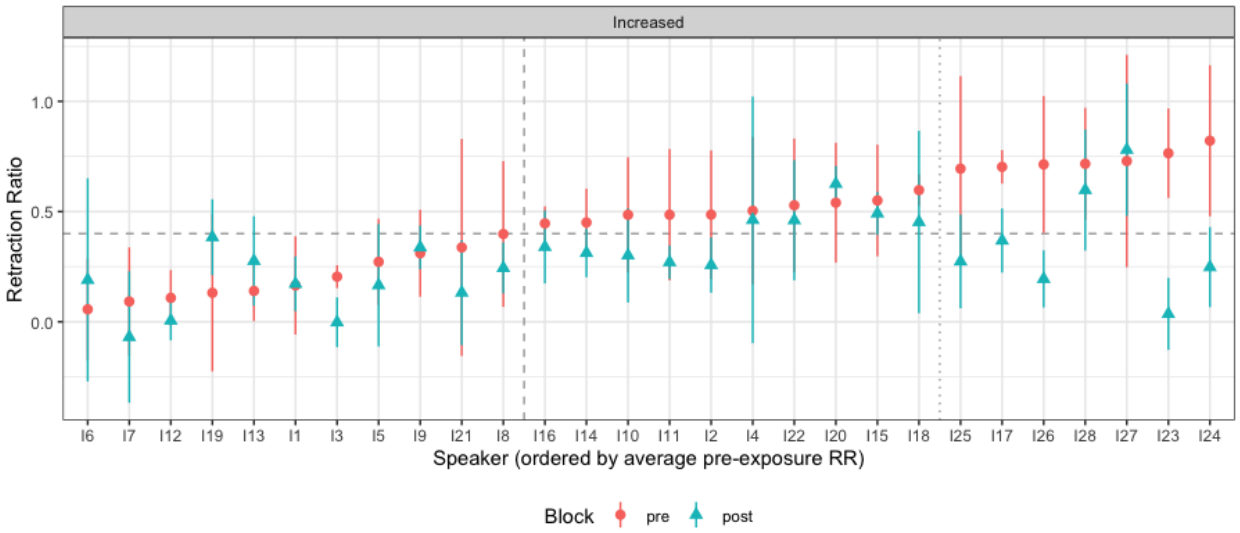


Figure 5.12: Mean and 95% confidence intervals of /skr/ pre- (red) and post-exposure (teal) retraction ratios for each speaker. The model talker's value is included for reference, indicated by the dashed gray lines intersecting both the x- and y-axes. The model talker's /str/ value (0.6) is indicated by a dotted gray line intersecting the x-axis.

patterns for /skr/ CLUSTERS for the same individuals, that is, individuals assigned to the increased RETRACTIONCONDITION. While many of the same individuals who have a baseline above the model talker in /str/ clusters in the increased RETRACTIONCONDITION also have a baseline well above the model in /skr/ clusters, there simply are more individuals with a baseline retraction ration above the model talker for /skr/ than /str/ clusters. Recall that the retraction conditions were defined in reference to the model talker's natural retraction ratios, and thus the increased retraction condition /str/ was modified to have a retraction ratio of 0.6 (indicated by a dotted gray line), while /skr/ was modified to have a retraction ratio of 0.4 (indicated by the dashed gray line). By pooling all individuals with a baseline retraction ratio above the model talker, we miss an important generalization. Individuals with a baseline /skr/ retraction ratio above 0.6, i.e. the individuals who produce a very /ʃ/-like /skr/, are likely to exhibit significant shifts toward the model talker (five out of seven: I17, 123, 124, I25, 126). However, individuals with a baseline /skr/ retraction ratio between 0.4 and 0.6, while still technically 'above' the model talker, are less likely to do so (zero out

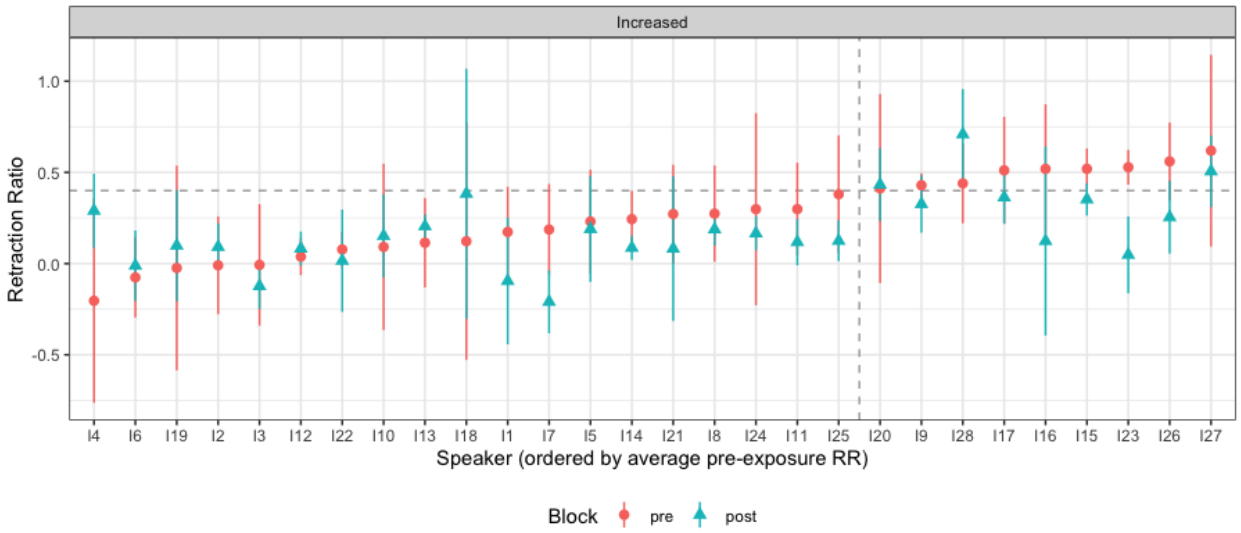


Figure 5.13: Mean and 95% confidence intervals of /spr/ pre- (red) and post-exposure (teal) retraction ratios for each speaker. The model talker's value is included for reference, indicated by the dashed gray lines intersecting both the x- and y-axes.

of seven).

Further supporting the observation that individuals above the model talker do not necessarily or automatically shift toward the model talker, /spr/ clusters show little evidence for convergence. Figure 5.13 plots the individual retraction ratio patterns for /spr/ Clusters, again for individuals assigned to the increased RetractionCondition. Here few significant shifts are observed for individuals with a baseline retraction ratio above the model talker, with only one out of nine participants (I23) exhibiting convergence. Importantly, no participant in the increased retraction condition, produced a baseline retraction ratio of 0.6 or higher for /spr/, which appeared to be the threshold for conditioning retraction in /str/ and /skr/ clusters.

Similarly, in the decreased retraction condition where the majority of participants have a baseline retraction ratio above the model talker, significant individual shifts are not the norm. Figure 5.14 returns to /str/ clusters, but this time for individuals assigned to the decreased RetractionCondition, where the model talker exhibited an /s/-like retraction ratio of 0.3 for /str/ clusters. As illustrated in Figure 5.14, only one out of eighteen individuals

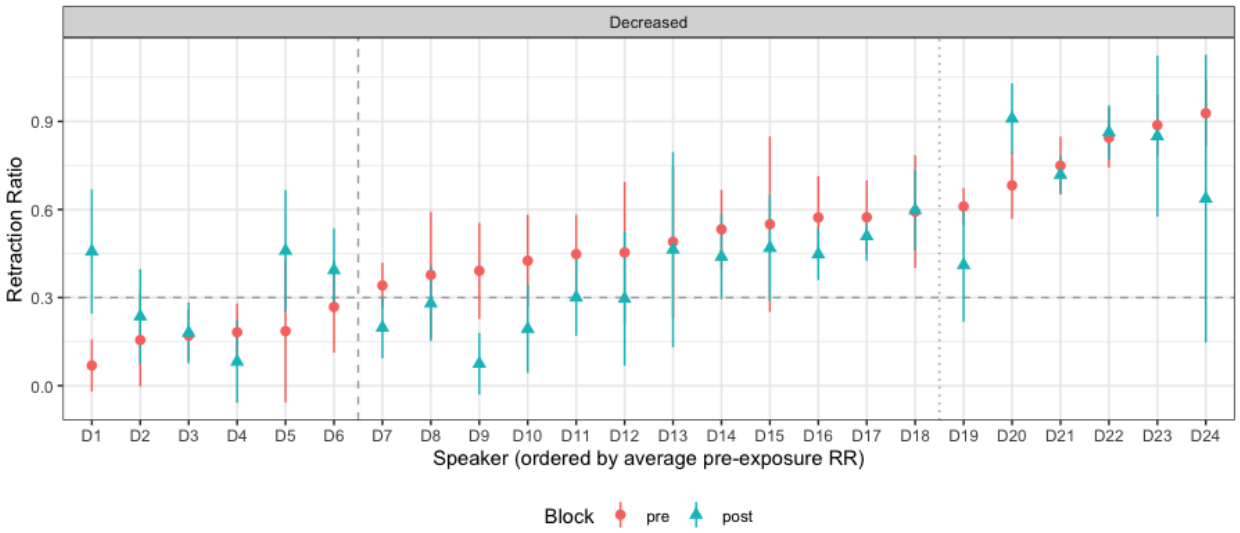


Figure 5.14: Mean and 95% confidence intervals of /str/ pre- (red) and post-exposure (teal) retraction ratios for each speaker. The model talker's value is included for reference, indicated by the dashed gray lines intersecting both the x- and y-axes. The model talker's /str/ increased value (0.6) is indicated by a dotted gray line intersecting the x-axis.

exhibits no overlap between their pre- and post-exposure confidence intervals. However, here it is an individual with a relatively low baseline retraction ratio (D9), in contrast to the increased retraction condition where the downward shifts come from the most retracted participants. However, fifteen out of eighteen individuals with a baseline above the model talker made shifts, whether significant or not, toward the model talker (D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D19, D21, D23, and D24), and out those with a baseline less than 0.6, twelve out of thirteen shifted toward the model talker. Furthermore, for individuals below the model talker, just one out of six (D1) exhibited a significant shift by increasing their relative degree of retraction, but four out of five (D1, D2, D3, D5, and D6) shifted in the direction of the model talker, and three of five (D1, D5, and D6) overshot the model talker. From these findings, it is apparent that individuals vary widely, but their shifts appear to be systematically condition by their baseline retraction ratios relative to the model talker.

5.5 Discussion

The present experiment asks if individuals converge toward increased and/or decreased degrees of /s/-retraction in /sCr/ clusters. The motivations behind this question are twofold. The first aim is to test the convergence path to sound change hypothesis in the laboratory, asking whether individuals can accelerate the trajectory of an ongoing sound change, at least for a moment, through convergence to a model talker. The second aim is to better understand the nature of the different constraints, including phonological, coarticulatory, and social, that condition when and where convergence is observed. In the present section, I discuss the findings of this experiment in light of these aims and the hypotheses laid out in Section 5.3.4.

First and foremost, this experiment asks if individuals exhibit convergence with respect to sibilants and specifically with respect to how preconsonantal sibilants are produced relative to the /s/-/ʃ/ contrast. The findings suggest that there is unequivocal empirical evidence that spontaneous convergence can be observed for /s/-retraction, although not universally. This contrasts with the findings of Kraljic et al. (2008) that convergence to /s/-retraction is only observed when participants are explicitly instructed to do so.

However, the observed convergence, found on three different metrics, by no means is ubiquitous or automatic. Firstly, I examine convergence through retraction ratio, which at its core asks if individuals exhibit differences in their relative production of the onset sibilant in /sCr/ between their pre- and post-exposure blocks. The first model illustrates that the retraction ratio lowers in the post-exposure block when the sibilant was manipulated to contain increased or hyper-increased retraction and when the participant exhibited a baseline above the model talker. This finding suggests that convergence is only observed toward a model talker exhibiting less retraction than the participant, with individuals producing a more /s/-like onset sibilant in response to a model talker with a more /s/-like onset sibilant. Secondly, the difference of distance metric asks if the relative distance in phonetic patterns of

retraction between the model talker and the participant changes between the pre- and post-exposure blocks. The difference in distance model suggests that on the whole individuals actually diverge from the model talker, and that convergence is only observed when the model talker exhibits less retraction than them. Finally, direction of shift asks whether individuals move in the direction of the model talker relative to their baseline, regardless of how big or smaller the actual shift is. The direction of shift model suggests that, in contrast to the previous model, convergence is the norm, and that while the more robust shifts are from individuals who produce a more retracted, i.e. more /ʃ/-like, onset sibilant than a model talker already producing a notably retracted sibilant, shifts by increasing an individual's degree of retraction are also consistently observed in the decreased retraction condition. The generalizations of the three different models paint a complicated picture for the hypotheses outlined earlier in this chapter. I'll work through the evidence for and against each hypothesis one at time.

Hypothesis 1 The first hypothesis states that listener-turned-speakers converge toward the model talker and do so regardless of the relative degree of retraction that the model talker exhibits. While the results of the three different models may be complicated, it is clear that the convergence patterns are moderated by the model talker's relative degree of retraction, characterized both by the retraction condition and the participant's baseline measurement relative to the model talker. This demonstrates unequivocally that Hypothesis 1 should be rejected.

Hypothesis 2 The second hypothesis states that the likelihood that a listener-turned-speaker exhibits convergence is constrained by the phonological contrast between /s/ and /ʃ/, even though such a contrast is not phonotactically relevant in the examined environments. Specifically, this predicts that convergence would only be observed toward decreased degrees of retraction in order to enhance the phonological contrast between /s/ and /ʃ/.

Convergence would not predicted toward increased degrees of retraction that would diminish that contrast. Such an observation would mirror Nielsen (2011), who observed convergence toward increased VOT, thus enhancing the voicing contrast, but not toward decreased VOT, which would diminish it. While the three models may differ in the predictions of what happens for individuals with a baseline below the model talker, all three models agree that more convergence is predicted for individuals above the model talker. Convergence is predicted in general by the retraction ratio model and the difference in distance model, and specifically in the increased and/or hyper-increased retraction condition, as predicted by the retraction ratio model and the direction of shift model. However, phonological contrast cannot explain why less convergence is observed in the decreased retraction condition where the sibilant contrast is the most enhanced, and it cannot explain why convergence is observed, albeit less frequently, for individuals with baseline below the model talker, especially in the decreased retraction condition. Taken together, the findings demonstrate that while the preservation of phonological contrast may play a critical role in predicting when and where convergence is observed, it cannot explain the whole picture.

Hypothesis 3a The third hypothesis contrasts with Hypothesis 2 and states that the likelihood that a listener-turned-speaker exhibits convergence is constrained by the coarticulatory naturalness of the shift. This predicts that convergence would be observed toward increased degrees of retraction, which by definition increase the degree of coarticulation, but not toward decreased degrees of retraction, which resist coarticulation. Such an observation would mirror Zellou et al. (2016), who observed convergence toward increased, but not decreased, nasalization. Mixed up with this, as /s/-retraction is a coarticulatory sound change by nature, increasing the degree of coarticulation also aligns the shift with the sound change in progress. Thus, this also predicts that convergence, as a social as well as linguistic and articulatory phenomenon, will be more likely if such a shift would align with the sound change than if it bucks the sound change. The findings of the present experiment do not

strongly support this hypothesis, as convergence is far more frequent and robust when an individual shifts to reduce their relative degree of retraction compared to when they shift to increase it. In fact, the only prediction from the three separate models that illustrates convergence by increasing coarticulation is from the direction of shift model, in which participants with a baseline below the model talker converge by increasing their retraction ratio, but only in the decreased, i.e. most /s/-like, retraction condition. In fact, no convergence is predicted for participants less retracted than the model talker in the increased retraction condition, and divergence is predicted for participants less retracted than the model talker in the hyper-increased retraction condition. While discouraging for the latter part of the present hypothesis about alignment with a sound change, as no convergence was observed precisely in the conditions where the model talker exemplifies the sound change, the observed pattern of convergence toward the model talker only in the decreased retraction condition may be evidence for convergence toward more natural coarticulatory patterns to the exclusion of the extreme instances that typify the sound change. Regardless, Hypothesis 3a cannot fully explain the findings of the present experiment, and, at best, can explain only one prediction of one of the collective models.

Hypothesis 3b The final hypothesis builds off of Hypothesis 3a to predict that, given both coarticulatory naturalness and alignment with a sound change, more convergence is predicted in /str/ clusters than /skr/ and /spr/ clusters. The question is largely moot, as Hypothesis 3a did not appear to adequately explain the findings. There is no Hypothesis 2b supported by Hypothesis 2, in which place of articulation is predicted to condition convergence toward decreased degrees of retraction, as the same phonological contrast between /s/ and /ʃ/ exists for all all /sCr/ clusters regardless of the intervening consonant. Nonetheless, turning to differences by cluster, the findings of the present experiment illustrate that more convergence is predicted in /str/ clusters relative to /skr/ and /spr/ clusters, and less robustly, /skr/ clusters relative to /spr/ clusters, and all of this is specifically in scenarios

by which convergence is realized by decreasing the retraction ratio. This finding may be unexpected, as if retraction is a phenomenon largely associated with /str/ clusters, why are individuals converging in those clusters precisely when they are producing less retraction? Firstly, it's possible that this is a result of the stimuli, in which the retraction ratio itself is different between /str/ clusters on the one hand and /spr/ and /skr/ on the other, such that the relatively more retracted sibilant in /str/ clusters induces more convergence. Secondly, it's possible that, despite no a priori predictions for a greater likelihood of converging toward a less retracted /str/ onset, these findings may be explained by the role of /str/ clusters signaling the sound change, encouraging potential convergence from individuals that identify with the model talker.

In sum, only Hypothesis 1 can be summarily rejected, and the remaining hypotheses each account for some of the findings but fail to capture the wealth of convergence patterns observed. Reflecting on the three models holistically and taking individual patterns into account, one generalization stands out: The most robust convergence is observed for individuals with a baseline above a substantially retracted model talker and individuals with a baseline below a non-retracted model talker. That is, in the increased (and when possible in the hyper-increased) retraction condition, a highly retracted participant will converge toward a retracted model talker, even though that means decreasing their relative degree of /s/-retraction. Likewise, in the decreased retraction condition, a non-retracted participant will converge toward a non-retracted model talker, even though that means increasing their relative degree of /s/-retraction. Elsewhere, when individuals are more retracted than a non-retracted model talker or less retracted than a retracted model talker (where the bulk of participants fall), different group patterns emerge depending on the measurement considered, but these hover around chance. One notable case is in the decreased retraction condition, where non-retracted individuals whose baseline is above the model talker, but not yet approaching the perceptual boundary between /s/ and /ʃ/, converge toward the model

talker by decreasing their relative degree of /s/-retraction. Here, non-retracted participants converge toward a non-retracted model talker by producing less retraction than their baseline. Collectively, these observations point to an in-group effect among the individuals at the edges of the community at large; retractors identify with and converge toward retractors, and non-retractors identify with and converge toward non-retractors. Participants converge toward /str/ clusters at higher rates, as /str/ clusters may hold a greater association with the sound change, and thus strengthen the identification between participant and model talker.

This in-group effect, by which retractors converge toward retractors and non-retractors converge toward non-retractors, may not have been predicted by the hypotheses focusing on the phonological and articulatory facets of convergence, but it is in line with the body of work understanding convergence as a social process. As Communication Accommodation Theory (Giles, 1973; Giles et al., 1973; Giles & Smith, 1979) asserts, convergence is a tool used by individuals to minimize social differences in order to signal solidarity, express agreement, or gain acceptance. In this vein, Communication Accommodation Theory can help explain some of the predictions based off social demographics, with queer speakers and speakers from the northeastern United States more likely to converge toward /s/-retraction than the general population. For Northeastern speakers, their increased likelihood to converge, predicted by the difference in distance and direction of shift models, may speak to a greater exposure to, and thus stronger indexical power of, /s/-retraction, as the sound change is generally more prevalent in the urban Northeast. For queer individuals, who are predicted in the retraction ratio model to decrease their retraction ratio in the post-test, in line with the most common observed patterns of convergence, this may speak less to an identification with the model talker, who is straight, and more to a greater desire to signal and gain acceptance.

With the present findings, I appeal to Communication Accommodation Theory not just to explain the differences between macro, self-reported demographic categories, but also to

interpret the role of the primary phonetic variable in question: /s/-retraction. This builds upon previous work in phonetic convergence that demonstrates that individuals appeal to their abstract linguistic knowledge when responding to an interlocutor: Nielsen (2008, 2011) shows that convergence is moderated by contrast maintenance; Babel (2009, 2012) and Sonderegger (2012) show that convergence is selectively observed for difference phonological features; Zellou et al. (2016) show that converge is influenced by coarticulatory factors. In the present experiment, the findings demonstrate that convergence is not separately mediated by social factors (like gender) and linguistic factors (like contrast enhancement), but rather by the interaction of the linguistic and the social: Listener-turn-speakers have detailed phonetic and phonological knowledge about both their own and their interlocutors' /sCr/ production relative to the prevocalic /s/ and /ʃ/ categories, determine whether they identify with their interlocutor based on that knowledge, and subsequently signal any potential identification with their interlocutor by selecting to shift their /sCr/ production.

Finally, returning to the proposal that convergence can be a mechanism for sound change actuation and propagation, the present experiment tests this hypothesis by asking whether a sound change can progress over the course of a brief interaction in the laboratory, but crucially does not speak to whether any such shifts persist or accumulate, as requisite under such a hypothesis. The present experiment challenges this account, or at the very least, suggests that, for /s/-retraction, convergence alone cannot have lead from a point of preliminary coarticulation to categorical sound change. This is because convergence is observed toward both increased and decreased retraction relative to a speaker's baseline for non-retractors and primarily only toward decreased retraction for retractors. In fact, the failure for retractors, i.e. individuals with a baseline above or near 0.6, to converge toward increased degrees of retraction, especially in the hyper-increased retraction condition, cannot be explained by the Communication Accommodation Theory. This suggests a potential role of contrast maintenance and challenges proposals for /s/-retraction propagation through convergence.

Furthermore, the in-group effect between both retractors and non-retractors reinforces the stratified distribution of individuals, with individuals firmly remaining within their predefined categories regardless of the manipulation of the model talker.

However, it is possible that given different experimental conditions, different generalizations may have emerged, potentially impacting the evidence for a convergence path to sound change. Specifically, it is possible that if a true continuum of retraction conditions were used, rather than the three (or initially two) used in the present experiment, that individuals along the retraction ratio continuum, including those near the group mean, would identify with and converge toward model talkers similar to them. Furthermore, it's possible that patterns may emerge such that convergence toward increased degrees of retraction are prevalent. However, at present, the current findings at the very least do not empirically support such a proposal, and at most, challenge that such a path exists for sound changes from below.

CHAPTER 6

CONCLUSIONS & IMPLICATIONS

In its final chapter, I draw this dissertation to a close, tying the threads back together and discussing the implications of the present findings. In Section 6.1, I summarize the results of the three experiments presented in the preceding chapters of this dissertation. In Section 6.2, I discuss what the findings of these experiments can tell us about sound change in general and /s/-retraction in particular, in light of the questions posed at the onset of this dissertation. Finally, in Section 6.3, I close this dissertation with a discussion of questions that arise from this work and the future research they motivate.

6.1 Summary of reported experiments

In this dissertation, I have reported the findings of three experiments that each focused on how individuals, as both listeners and speakers, make use of their detailed phonological knowledge about /s/-retraction in their response to an interlocutor exhibiting a prescribed degree of retraction. With these experiments, I asked how the different patterns in perception and production observed for a sound change in progress shed light on both the origins and future of the sound change and how sound change transitions and propagates generally.

Experiment I: Cue Integration The first experiment reported in this dissertation examines lexical identification of words with /sCr/ onsets to ask if and when listeners use /s/-retraction in speech processing. Using eye tracking in the Visual World Paradigm, participants were presented with two images, one representing a word with an /sC/ onset, like *sting*, and one representing a word with an /sCr/ onset, like *string*. Participants were then given oral instructions to select one of the images, with the auditory stimuli containing manipulated degrees of /s/-retraction. With the high temporal resolution of eye tracking, it is possible to know exactly what a listener was hearing when they planned and executed an

eye movement. With this design, this experiment asked if listeners are able to use the cues of /s/-retraction to accurately predict the presence of an upcoming /r/, correctly looking to a word like *string* before they even hear the disambiguating /r/.

The results of this experiment demonstrated that on the whole most listeners wait until well after the onset of the disambiguating /r/ to look toward the correct image on the screen. However, the results also definitively demonstrated that listeners can and do use those cues, just not as often as they wait. Early looks to the correct image are most common when the cues of /s/-retraction are the strongest and when they are most expected, as in /str/ clusters. These findings demonstrated that listeners can immediately integrate the spectral cues of retraction when considering the different lexical candidates, even if they don't make the ultimate decision faster or more accurately.

Experiment II: Categorization The second experiment reported in this dissertation examined the categorization of onset sibilants to ask if listeners account for their potential experience with /s/-retraction by adjusting their phonemic boundaries in the environments where it's most expected. Participants were presented with a word beginning with any step on a continuum from /s/ and /ʃ/ and asked to select the orthographic representation for the word that they heard. However, due to the phonotactic restrictions of English, by which only /ʃ/ is observed preceding /r/ and only /s/ preceding all other consonants, it was necessary to use nonce words in order to force a contrast between /s/ and /ʃ/ preconsonantally.

First and foremost, the results demonstrated that listeners are much less categorical in their perception of preconsonantal sibilants than prevocalic sibilants, doubtless because of the lack of native phonological contrast in those environments. The results further demonstrated that individuals are less categorical in /str/ clusters than /spr/ and /skr/. This observation is akin to perceptual compensation for coarticulation, but manifests itself as a dampened categorization response curve rather than the traditional shifted categorization response curve due to the lack of phonological contrast. Whether we want to call this compensation

or decreased categoricity, the findings of this experiment demonstrated that individuals shift their perceptual strategies for a sound change in progress as a result of their experience and expectations.

Due to the simplicity of the experimental design, it was possible to run Experiment II online in order to attract a more diverse participant pool. Specifically, with participants of varying ages, I asked how categorization strategies for /s/-retraction are changing in apparent time. The findings of this experiment demonstrated that younger listeners are less categorical in their perception of preconsonantal sibilants than older listeners, adjusting their perceptual strategies as a result of their increased experience with, and thus increased expectation for, /s/-retraction. Interestingly, younger listeners are less categorical than older listeners not just for /str/ clusters, but also for /spr/ and /skr/ clusters, highlighting that these environments may be the next loci for the change, albeit in very incipient stages.

Finally, Experiment II asked how different measurements of masculinity, and of masculine toughness specifically, influence listeners' categorization strategies in order to shed light on the potential socio-indexicality of the sound change. The findings of this experiment suggest that preconsonantal sibilants index masculinity much less strongly than prevocalic sibilants, and /str/ clusters do so particularly weakly. This suggests that /s/-retraction as a sound change is not primarily about performing or indexing masculinity, but may do other socio-indexical work not examined here.

Experiment III: Convergence The third experiment reported in this dissertation asked what happens when the listener takes a turn as a speaker. Specifically, this experiment employed a covert shadowing task, meaning that participants were never explicitly instructed to imitate the model talker. This experiment asked how an individual may shift their own relative /s/-retraction as a result of exposure to a model talker manipulated to exhibit varying degrees of /s/-retraction. At stake here is whether /s/-retraction as a sound change in progress can be 'accelerated' in the laboratory or whether factors like the maintenance of

the phonological contrast between /s/ and /ʃ/ hinder convergence toward /s/-retraction.

The results of this experiment demonstrated that convergence of sibilant spectral cues is observed, but that it is by no means automatic or universal. Rather, in characterizing convergence using three different metrics, a complicated picture appeared for exactly under which conditions convergence is predicted. On the whole, a general pattern emerged that suggests that like converges to like. Individuals who come into lab already very retracted, i.e. /ʃ/-like, will converge toward a more retracted model talker, even if that means reducing their relative degree of retraction and sounding more /s/-like. On the other hand, individuals who come into lab producing little to no retraction at all will converge toward a less retracted, i.e. more /s/-like, model talker, even if that means increasing their relative degree of retraction and sounding more /ʃ/-like. However, there was one notable gap in this paradigm: I did not find evidence to suggest that a very retracted, i.e. /ʃ/-like, individual will converge toward a model talker more retracted than them, thereby increasing their degree of retraction and sounding more /ʃ/-like. These findings suggest that convergence alone cannot be a mechanism of sound change transition and propagation, as /s/-retraction was not accelerated through the shadowing task.

6.2 General discussion & conclusions

In the three experiment reported in this dissertation, I have provided a window into how listeners and speakers perceive and produce a sound change in progress in order to better understand the process by which sound change transitions and propagates. I selected /s/-retraction as a case study for this dissertation as it has the unique profile of being a sound change in progress not tied to a specific region, community, or demographic.

In both Experiments I and II, I have empirically demonstrated that listeners in fact have detailed phonological knowledge about /s/-retraction. That phonological knowledge not only concerns the presence of an upcoming /r/ but also the phonological nature of the intervening

consonant. Given that listeners have an understanding of this context-dependent variation, the question was what do they do with that knowledge? Experiment I: Cue Integration used eye-tracking during a lexical identification task and empirically demonstrated that listeners can immediately use the cues of /s/-retraction in speech processing, and do so in the environments where those cues are most available and most expected. Experiment II: Categorization used a phoneme categorization task with nonce words and empirically demonstrated that listeners are less categorical in their perception of onset sibilants in the environments in which /s/-retraction is most expected and that categoricity is changing over time, as generations have different experiences with and thus different expectations for /s/-retraction. Experiment III: Convergence used a covert shadowing task and empirically demonstrated that individuals can converge toward the spectral cues of /s/-retraction, but that individuals are most likely to converge toward an interlocutor with a similar phonological pattern to their own. That is, like converges to like.

One notable finding of this dissertation was simply that the cues of retraction are available to listeners. And not only are such cues available, but listeners make use of them in real time. Previous work had found that listeners appear to wait until the onset of a vowel before using spectral cues from sibilants in speech processing and hypothesized that this apparent 'buffer' strategy may be a consequence of the high degree of context-dependent variation exhibited by sibilants (Galle et al., 2019). However, this dissertation not only found that listeners can immediately use the spectral cues for sibilants, but that they can do so specifically for the cues of that extact context-dependent variation. These findings suggest that, for some individuals, /s/-retraction is an important cue for the presence of the upcoming /r/, in part because it's simply available earlier. While the results firmly show that we are no where near the stage at which such cues become more important than the actual /r/ in contrasting words like *string* and *sting*, it's not impossible to imagine such a future where the /r/ becomes redundant and that contrast shifts to being between words

like *shting* and *sting*. That is to say, the present findings suggest that, during the transition of the sound change, the anticipatory cues of /s/-retraction may continue to be weighted more strongly while the later /r/ may shift to be weighted less strongly.

Another notable finding of this dissertation is that while convergence was empirically demonstrated for sibilants, it was not universal nor was it consistently conditioned by coarticulatory naturalness or maintenance of a phonological contrast. This may be in part because these factors, which have previously been hypothesized to condition convergence (e.g. Nielsen, 2011; Zellou et al., 2016), are in direct contrast when it comes to /s/-retraction: Increasing the degree of coarticulation by definition diminishes the phonological contrast between /s/ and /ʃ/ (so long as such a contrast is presumed to exist). Rather, convergence is most frequently observed toward an interlocutor who exhibits a pattern similar to the participant. This finding supports our understanding of convergence as a social process and a social tool to signal acceptance and connection, strengthening sociological theories like Communication Accommodation Theory (Giles, 1973, i.a.). Crucially, however, the convergence task failed to accelerate the sound change in the laboratory, as participants did not robustly or consistently increase their degree of retraction. These findings demonstrated that convergence alone cannot have led to the current distribution of /s/-retraction nor can they lead us from the current state to categorical, community-wide change. Much of the work appealing to a convergence path to sound change is concerned with change in the context of dialect leveling and dialect contact (e.g. Trudgill, 1981) and perhaps in those contexts a convergence path to sound change remains feasible. However, the empirical findings of this dissertation challenge that such a path exists for a sound change from below, like /s/-retraction.

One surprising finding of this work was the apparent role that the phonological contrast between /s/ and /ʃ/ plays in preconsonantal environments. Recall that despite the contrast between /s/ and /ʃ/ made prevocalically, the sibilants are in complementary distribution preconsonantally: Only /ʃ/ precedes /r/, while /s/ precedes all other consonants. Less

surprising was the role this played in categorization. Listeners were beyond a doubt less categorical than would be expected prevocalically, even sometimes hearing the last step on the continuum, comprised of 100% prevocalic /ʃ/, as /s/ in /str/ clusters. However, they generally continued to make a contrast when encouraged to do with nonce word orthography. However, the role of contrast maintenance was particularly surprising in convergence. As I mentioned in the preceding paragraph, contrast maintenance cannot consistently predict convergence, meaning that listeners do not consistently converge toward less retracted, i.e. more /s/-like, /sCr/ onsets. However, when participants do exhibit convergence, that is when they converge toward someone exhibiting a similar pattern to them, they are significantly more likely to decrease their relative degree of retraction than increase it. This pattern is even stronger for retractors than non-retractors. This suggests that while the shift itself may be driven by identification with an interlocutor, the direction of the shift may be constrained by phonological contrast. Thus, convergence by retractors appears to be constrained by a phonological contrast both not maintained phonotactically and not maintained phonetically by either the speaker or their interlocutor. This puzzling observation highlights the complicated interplay between social information and phonological structures in convergence in particular and the link between speech perception and production more generally.

While /s/-retraction has received significant attention from a production standpoint, little work has previously examined perception. This dissertation makes a valuable contribution by providing empirical evidence of perception for /s/-retraction, including both online and offline experiments. Even the convergence data, while analyzing produced speech, relies heavily on what was perceived by the listener-turned-speaker. This dissertation additionally breaks from previous research by including an examination of /spr/ and /skr/ clusters as potential environments for coarticulatory /s/-retraction. From the results reported in this dissertation, it is clear that /s/-retraction is in fact most advanced in /str/ clusters in production. Similarly, from the perceptual experiments, listeners expect and account for

/s/-retraction the most in those clusters. However, significant retraction is observed in both /spr/ and /skr/ clusters and listeners do use the cues of the retraction in those environments as well. In fact, in terms of baseline retraction ratios and convergence to increased degrees of retraction, /str/ and /skr/ clusters patterned more similarly than /spr/ and /skr/ clusters. Thus, to define /s/-retraction as a phenomenon limited to /str/ clusters alone is to create a false dichotomy between the three places of articulation. Furthermore, the observation that the categorization strategies of all clusters are moving consistently in apparent time, taken together with the variation in produced retraction ratios, suggest that /spr/ and /skr/ are the next loci for the sound change. Eventually, they may begin to cross the perceptual boundary between /s/ and /ʃ/, such that *script* may really sound like *shcript*.

6.3 Limitations & future directions

One methodological limitation of this research was that the same participants were not recruited for all three experiments. Such a design would have allowed for an examination of how individual patterns in the perception experiments may condition or align with different individual patterns in production. For instance, it is currently unknown whether the individuals who robustly use the spectral cues of /s/-retraction as soon as they are available are also the individuals who are more likely to shift their production in response to their interlocutor. Furthermore, the production patterns for participants in Experiments I and II is unknown, such that we cannot say with any certainty whether individuals who produce more retraction to begin with are more likely to use those cues in perception. This limitation was a consequence of the robust individual variation observed in the three experiments, which necessitated recruiting more subjects and creating additional conditions in order to get a better picture of how (or if) that variation was structured. After the inclusion of more participants and the conditions, it was no longer logistically feasible to rerun the experiments together with the same subject pool. Future work could do so in order to better understand

the relationship between when cues are immediately integrated and when those cues are converged to.

This dissertation found no evidence for a convergence path to sound change: Individuals did not converge toward increased degrees of /s/-retraction robustly, which would have accelerated the sound change in the laboratory. Future research may expand upon this finding to examine convergence to a sound change in progress longitudinally. Such a longitudinal approach was not feasible as a dissertation research project, but as part of a larger, longitudinal project may shed light on how small conversational shifts may persist and accumulate, well past the single session in the laboratory. Particularly, it would be advantageous to examine other sound changes in progress, including changes with well-documented social meaning and changes that may not be constrained by existing phonological contrasts, in order to better test convergence as a mechanism for sound change actuation, propagation, and transition.

APPENDIX A

A.1 Statistical models with all main effects and interactions

In this appendix, the predictions of the mixed effects models for Experiments I, II, and II are provided in full, including effects and interactions that did not reach the significance threshold of 0.05.

Experiment I: Cue Integration (Chapter 3)

Table A.1: Complete model predictions for prevocalic sibilants in Experiment I: Cue Integration, including all main effects and interactions. This model predicts fixation accuracy for prevocalic /s/ vs. /ʃ/, N=26205. A positive value indicates a greater prediction of fixations on the target word.

	Est.	*SE*	*t*	*p*
(Intercept)	0.88	1.07	0.83	0.41
TrialOrder	0.05	0.01	3.45	< 0.001
TimeWindow	0.27	0.04	7.31	< 0.001
OnsetSH	−0.32	0.20	−1.61	0.11
Increased	−0.40	0.50	−0.81	0.42
Hyper	−0.62	0.43	−1.43	0.16
TimeWindow:OnsetSH	−0.03	0.05	−0.65	0.51
TimeWindow:Increased	0.19	0.05	3.79	< 0.001
TimeWindow:Hyper	0.12	0.05	2.63	0.002
OnsetSH:Increased	0.19	0.28	0.69	0.49
OnsetSH:Hyper	0.24	0.24	0.99	0.32
TimeWindow:OnsetSH:Increased	0.14	0.07	1.91	0.06
TimeWindow:OnsetSH:Hyper	0.06	0.07	0.93	0.34
Male	−0.35	0.29	−1.24	0.22
Queer	−0.25	0.29	−0.84	0.40
Northeast	−0.51	0.33	−1.53	0.12
South	0.23	0.37	0.62	0.53
West	−0.64	0.34	−1.89	0.05
Empathy	−0.01	0.01	−0.63	0.53
Toughness	0.02	0.13	0.18	0.86
Openness	0.004	0.3	0.16	0.87
Anxiety	−0.01	0.01	−1.63	0.11

Table A.2: Complete model predictions for preconsonantal sibilants in Experiment I: Cue Integration, including all main effects and interactions. This model predicts fixation accuracy for /sC/ vs. /sCr/ clusters in different phonological environments with differing cues of retraction, N=26415. A positive value indicates a greater prediction of fixations on the target word.

	Est.	*SE*	*t*	*p*
(Intercept)	−0.17	0.45	−0.39	0.700
TrialOrder	0.03	0.08	0.41	0.679
TimeWindow	0.11	0.03	2.89	0.003
SCR	−0.23	0.05	−4.47	< 0.001
Place1	−0.03	0.08	−0.38	0.703
Place2	0.10	0.25	0.40	0.686
Increased	−1.59	0.63	−2.52	0.016
Hyper	−0.92	0.36	−2.52	0.016
TimeWindow:SCR	−0.03	0.05	−0.65	0.511
TimeWindow:Place1	−0.04	0.06	−0.69	0.487
TimeWindow:Place2	0.08	0.07	1.09	0.272
TimeWindow:Increased	0.21	0.05	3.72	< 0.001
TimeWindow:Hyper	0.09	0.04	1.99	0.041
SCR:Place1	0.10	0.11	0.92	0.357
SCR:Place2	−0.31	0.13	−2.40	0.017
SCR:Increased	0.34	0.07	4.40	< 0.001
SCR:Hyper	0.43	0.07	6.33	< 0.001
Increased:Place1	0.20	0.11	1.77	0.077
Increased:Place2	−0.22	0.37	−0.62	0.532
Hyper:Place1	0.20	0.10	1.97	0.047
Hyper:Place2	0.13	0.32	0.40	0.684
TimeWindow:SCR:Place1	0.19	0.06	3.12	0.002
TimeWindow:SCR:Place2	−0.01	0.06	−0.10	0.917
TimeWindow:SCR:Increased	−0.06	0.08	−0.82	0.410
TimeWindow:SCR:Hyper	0.04	0.06	0.54	0.592
TimeWindow:Increased:Place1	0.06	0.08	0.69	0.491
TimeWindow:Increased:Place2	−0.21	0.18	−1.13	0.273
TimeWindow:Hyper:Place1	0.03	0.07	0.40	0.687
TimeWindow:Hyper:Place2	−0.02	0.08	−0.19	0.852
SCR:Increased:Place1	0.13	0.14	0.91	0.364
SCR:Increased:Place2	0.21	0.19	1.11	0.263
SCR:Hyper:Place1	0.49	0.17	3.04	0.002
SCR:Hyper:Place2	0.18	0.16	1.12	0.263

Table A.2 continued

	Est.	*SE*	*t*	*p*
Male	−0.49	0.32	−1.55	0.127
Queer	0.001	0.33	0.01	0.995
Northeast	−0.41	0.37	−1.11	0.264
South	0.51	0.40	1.25	0.216
West	−0.57	0.37	−1.53	0.124
Empathy	−0.27	0.15	−1.83	0.075
Toughness	0.08	0.17	0.49	0.627
Openness	0.08	0.15	0.52	0.604
Anxiety	−0.13	0.14	−0.97	0.333

Experiment II: Categorization (Chapter 4)

Table A.3: Complete model predictions for Experiment I: Categorization, including all main effects and interactions. This model predicts sibilant categorization in different phonological environments and with different indicators of masculine stereotypes of toughness, N=18476. A positive value indicates stronger /ʃ/ prediction.

	Est.	*SE*	*z*	*p*
(Intercept)	−0.69	0.07	−10.72	< 0.001
Trial Order	0.09	0.14	0.65	0.51
Step	1.00	0.02	41.03	< 0.001
Age	0.06	0.07	0.89	0.37
Cluster1	0.14	0.09	1.51	0.13
Cluster2	0.20	0.09	2.31	< 0.05
ToughnessEndorsement	0.09	0.07	1.36	0.17
TalkerToughness	0.04	0.04	0.95	0.34
FaceToughness	0.09	0.04	2.28	< 0.05
Step:Age	0.19	0.02	7.82	< 0.001
Step:Cluster1	0.19	0.02	7.79	< 0.001
Step:Cluster2	0.11	0.01	7.94	< 0.001
Step:ToughnessEndorsement	−0.09	0.02	−4.07	< 0.001
Step:TalkerToughness	−0.02	0.02	−0.86	0.39
Step:FaceToughness	−0.05	0.02	−2.24	< 0.05
Age:Toughness	0.13	0.07	1.87	0.06
Cluster1:ToughnessEndorsement	0.003	0.03	0.10	0.92
Cluster2:ToughnessEndorsement	−0.02	0.02	1.00	0.32
Cluster1:TalkerToughness	0.09	0.02	2.67	< 0.001
Cluster2:TalkerToughness	−0.01	0.02	−0.67	0.49

Table A.3 continued

	Est.	*SE*	*t*	*p*
Cluster1:FaceToughness	0.04	0.03	1.59	−0.11
Cluster2:FaceToughness	−0.01	0.02	−0.83	0.40
Step:Cluster1:ToughnessEndorsement	−0.05	0.02	−1.99	< 0.05
Step:Cluster2:ToughnessEndorsement	−0.03	0.01	−1.76	0.08
Step:Cluster1:TalkerToughness	−0.04	0.02	−1.41	0.16
Step:Cluster2:TalkerToughness	0.01	0.01	0.37	0.71
Step:Cluster1:FaceToughness	0.001	0.02	0.05	0.92
Step:Cluster2:FaceToughness	0.001	0.02	0.06	0.95
Step:Age:ToughnessEndorsement	0.12	0.03	4.85	< 0.001

Experiment III: Convergence (Chapter 5)

Table A.4: Complete model predictions for retraction ratio in Experiment III: Convergence, including all main effects and interactions. This model predicts changes in retraction as a result of exposure to a model talker with varied degrees of retraction, N=2854. A positive value indicates a higher, i.e. more /ʃ/ like, retraction ratio.

	Est.	*SE*	*t*	*p*
(Intercept)	0.09	0.04	1.82	0.069
Increased	0.17	0.05	3.03	0.002
Hyper	0.30	0.06	5.45	< 0.001
Cluster1	0.19	0.05	3.82	< 0.001
Cluster2	−0.002	0.07	−0.04	0.970
Post	0.19	0.04	5.21	< 0.001
Above	0.32	0.03	9.75	< 0.001
Increased:Cluster1	−0.05	0.05	−0.86	0.391
Increased:Cluster2	0.07	0.07	1.09	0.278
Hyper:Cluster1	0.001	0.05	0.03	0.978
HyperCluster2	0.09	0.07	1.29	0.196
Increased:Post	−0.17	0.04	−3.97	< 0.001
Hyper:Post	−0.18	0.04	−4.52	< 0.001
Cluster1:Post	0.02	0.05	0.31	0.758
Cluster2:Post	0.07	0.07	1.05	0.293
Increased:Above	0.03	0.05	0.69	0.490
Hyper:Above	−0.07	0.09	−0.74	0.457
Cluster1:Above	0.07	0.05	1.30	0.192
Cluster2:Above	0.09	0.06	1.46	0.145
Post:Above	−0.17	0.04	−4.34	< 0.001

Table A.4 continued

	Est.	*SE*	*t*	*p*
Increased:Cluster1:Post	−0.03	0.05	−0.65	0.515
Increased:Cluster2:Post	−0.04	0.06	−0.63	0.532
Hyper:Cluster1:Post	0.01	0.06	0.14	0.886
Hyper:Cluster2:Post	−0.09	0.08	−1.13	0.259
Increased:Cluster1:Above	0.11	0.06	1.84	0.065
Increased:Cluster2:Above	−0.02	0.07	−0.24	0.813
Increased:Post:Above	−0.05	0.06	−0.96	0.339
Hyper:Post:Above	−0.01	0.11	−0.07	0.942
Cluster1:Post:Above	−0.12	0.05	−2.23	0.025
Cluster2:Post:Above	−0.13	0.07	−2.04	0.041
Male	0.01	0.02	0.29	0.773
Queer	−0.01	0.04	−0.13	0.898
Northeast	−0.02	0.05	−0.50	0.616
South	0.04	0.05	0.72	0.471
West	0.02	0.06	0.35	0.727
Empathy	0.02	0.02	0.88	0.381
MRAS	−0.01	0.02	−0.37	0.712
Openness	−0.01	0.02	−0.48	0.629
Anxiety	−0.02	0.02	−0.97	0.331
Post:Male	−0.02	0.01	−1.78	0.075
Post:Queer	−0.08	0.02	−3.68	< 0.001
Post:Northeast	−0.03	0.02	−1.41	0.159
Post:South	−0.05	0.03	−1.67	0.095
Post:West	−0.01	0.03	−0.46	0.644
Post:Empathy	−0.01	0.01	−0.68	0.498
Post:MRAS	−0.04	0.01	−3.40	< 0.001
Post:Openness	0.004	0.01	0.41	0.681
Post:Anxiety	0.04	0.01	4.18	< 0.001

Table A.5: Complete model predictions for difference in distance in Experiment III: Convergence, including all main effects and interactions. This model predicts changes in difference in distance as a result of exposure to a model talker with varied degrees of retraction, N=816. A positive value indicates greater convergence toward the model talker.

	Est.	*SE*	*z*	*p*
(Intercept)	−0.07	0.03	−1.97	0.049
Increased	0.03	0.04	0.88	0.378
Hyper	−0.01	0.04	−0.29	0.768
Cluster1	0.03	0.05	0.67	0.505
Cluster2	−0.03	0.07	−0.51	0.611
Above	0.07	0.03	2.08	0.037
Increased:Cluster1	−0.02	0.07	−0.29	0.773
Increased:Cluster2	0.05	0.08	0.61	0.540
Hyper:Cluster1	−0.04	0.06	−0.56	0.577
Hyper:Cluster2	−0.01	0.07	−0.13	0.893
Increased:Above	−0.06	0.04	−1.40	0.162
Hyper:Above	−0.13	0.05	−2.19	0.028
Cluster1:Above	−0.01	0.06	−0.10	0.923
Cluster2:Above	0.08	0.07	1.11	0.266
Increased:Cluster1:Above	0.01	0.08	0.18	0.860
Increased:Cluster2:Above	0.05	0.09	0.52	0.606
Hyper:Cluster1:Above	−0.24	0.12	−2.09	0.036
Hyper:Cluster2:Above	0.06	0.13	0.47	0.639
Male	−0.003	0.01	−0.27	0.787
Queer	0.002	0.02	0.09	0.930
Northeast	0.05	0.02	1.99	0.045
South	−0.03	0.03	−0.91	0.363
West	0.03	0.03	0.90	0.366
Empathy	0.01	0.01	1.03	0.304
MRAS	−0.02	0.01	−1.61	0.107
Openness	−0.01	0.01	−0.86	0.389
Anxiety	0.02	0.01	1.37	0.171
Increased:Above	0.03	0.05	0.69	0.490
Hyper:Above	−0.07	0.09	−0.74	0.457

Table A.6: Complete model predictions for direction of shift in Experiment III: Convergence, including all main effects and interactions. This model predicts changes in direction of shift as a result of exposure to a model talker with varied degrees of retraction, N=816. A positive value indicates greater convergence toward the model talker.

	Est.	*SE*	*z*	*p*
(Intercept)	0.86	0.41	2.31	0.021
Increased	−1.48	0.47	−3.09	0.002
Hyper	−1.86	0.47	−3.90	< 0.001
Cluster1	−1.30	0.74	−1.77	0.077
Cluster2	0.43	1.05	0.41	0.683
Above	−0.89	0.41	−2.15	0.031
Increased:Cluster1	1.42	0.75	1.90	0.050
Increased:Cluster2	−1.86	1.17	−1.58	0.114
Hyper:Cluster1	1.59	0.73	2.16	0.031
Hyper:Cluster2	−1.02	1.16	−0.88	0.376
Increased:Above	3.34	0.60	5.55	< 0.001
Hyper:Above	7.65	711.21	0.01	0.991
Cluster1:Above	1.52	0.74	2.05	0.040
Cluster2:Above	−0.53	1.12	−0.47	0.637
Male	−0.14	0.14	−1.04	0.299
Queer	0.11	0.31	0.36	0.718
Northeast	0.99	0.31	3.13	0.002
South	−0.10	0.37	−0.27	0.787
West	0.86	0.38	2.25	0.024
Empathy	0.10	0.15	0.70	0.482
MRAS	−0.15	0.14	−1.05	0.295
Openness	−0.17	0.15	−1.18	0.238
Anxiety	0.05	0.14	0.34	0.732

A.2 Stimuli materials

In this appendix, I provide the complete list of stimuli for Experiments I: Cue Integration (Chapter 3) and III: Imitation (Chapter 5). Each item is indicated as to whether it was included as an audiovisual stimulus in Experiment I, an auditory stimulus in Experiment III, and/or a reading stimulus in Experiment III.

Table A.7: Complete list of stimuli for Experiment I and III

word	IPA	SUBTL_{WF}	Experiment I	Experiment III
big	/bɪg/	682.82	audiovisual	reading and listening
bin	/bɪn/	5.37		reading only
brew	/bru/	2.51		reading and listening
brick	/brɪk/	10.18	audiovisual	reading and listening
brim	/brɪm/	0.88		reading only
bruise	/bruz/	3.24		reading only
boom	/bum/	21.80		reading only
boot	/but/	11.14		reading and listening
chute	/ʃut/	3.61		reading only
coo	/ku/	0.69		reading only
coop	/kup/	10.35		reading and listening
crew	/kru/	47.53		reading and listening
crimp	/krɪmp/	0.43		reading only
crude	/krud/	3.04		reading only
crypt	/krɪpt/	1.37	audiovisual	reading and listening
dew	/du/	2.14		reading and listening
din	/dɪn/	1.18		reading only
dip	/dɪp/	7.96	audiovisual	reading and listening
drew	/dru/	25.04		reading and listening
drink	/drɪŋk/	247.39		reading only
drip	/drɪp/	5.12	audiovisual	reading and listening
drool	/drul/	2.16		reading only
dune	/dun/	1.00		reading only
gift	/gɪft/	64.51	audiovisual	reading and listening
gill	/gɪl/	1.71		reading only
goof	/guf/	2.22		reading only
goop	/gup/	0.69		reading and listening
grill	/grɪl/	4.45		reading only
grip	/grɪp/	9.69	audiovisual	reading and listening
groove	/gruv/	4.16		reading only
group	/grup/	73.76		reading and listening
kick	/kɪk/	73.41		reading only
kit	/kɪt/	17.65	audiovisual	reading and listening

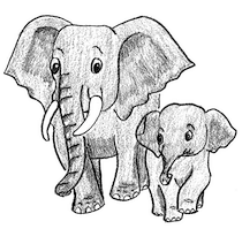
Table A.7 continued

word	IPA	$SUBTL_{WF}$	Experiment I	Experiment III
pick	/pɪk/	198.39		reading only
pig	/pɪg/	39.14	audiovisual	reading and listening
pooch	/putʃ/	1.29		reading and listening
pool	/pul/	46.98		reading only
prick	/prɪk/	14.12	audiovisual	reading and listening
prim	/prɪm/	0.37		reading only
prove	/pruv/	70.39		reading only
prune	/prun/	1.47		reading and listening
scoop	/skup/	5.67		reading and listening
scoot	/skut/	2.1847		reading only
screw	/skru/	37.49		reading and listening
scrooge	/skrudʒ/	3.86		reading only
scribble	/skrɪbəl/	0.63		reading only
script	/skrɪpt/	19.61	audiovisual	reading and listening
shift	/ʃɪft/	22.82		reading only
ship	/ʃɪp/	98.88	audiovisual	reading and listening
shin	/ʃɪn/	3.08		reading only
shit	/ʃɪt/	474.65	audiovisual	reading and listening
shoe	/ʃu/	30.39		reading and listening
shoes	/ʃuz/	30.39		reading only
shoot	/ʃut/	164.94		reading and listening
sick	/sɪk/	165.43		reading only
sift	/sɪft/	0.75		reading only
sip	/sɪp/	5.10	audiovisual	reading and listening
sit	/sɪt/	311.35	audiovisual	reading and listening
skin	/skɪn/	44.04		reading only
skip	/skɪp/	21.1	audiovisual	reading and listening
soon	/sun/	257.65		reading only
soothe	/suð/	1.29		reading only
spin	/spɪn/	14.63		reading only
spit	/spɪt/	19.35	audiovisual	reading and listening
spool	/spul/	0.51		reading only
spoon	/spun/	7.61		reading and listening
spring	/sprɪŋ/	31.31		reading only
spritz	/sprɪts/	0.49	audiovisual	reading and listening
spruce	/sprus/	1.1		reading and listening
sprue	/spru/	0.00		reading only
stew	/stu/	6.43		reading and listening
sting	/stɪŋ/	7.02	audiovisual	reading and listening
stint	/stɪnt/	0.75		reading only

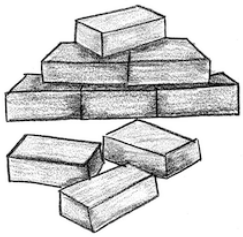
Table A.7 continued

word	IPA	$SUBTL_{WF}$	Experiment I	Experiment III
stool	/stul/	3.51		reading only
strewn	/strun/	0.37		reading and listening
string	/strɪŋ/	12.67	audiovisual	reading and listening
strip	/strɪp/	15.69		reading only
strudel	/struɾəl/	0.92		reading only
sue	/su/	29.37		reading and listening
suit	/sut/	68.61		reading and listening
tin	/tɪn/	8.65		reading only
tip	/tɪp/	27.63	audiovisual	reading and listening
trim	/trɪm/	4.27		reading only
trip	/trɪp/	82.39	audiovisual	reading and listening
true	/tru/	253.35		reading and listening
truth	/truθ/	192.18		reading only
tune	/tun/	15.61		reading only
two	/tu/	1066.35		reading only

In this appendix, the visual stimuli used for Experiment I: Cue Integration (Chapter 3) are provided, listed in alphabetical order with their corresponding labels.



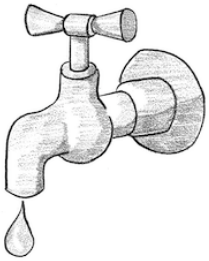
big



brick

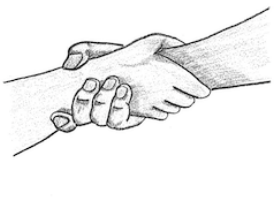
crypt

dip



drip

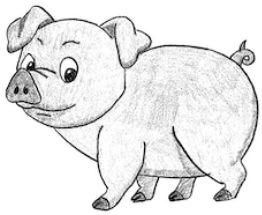
gift



grip

kit

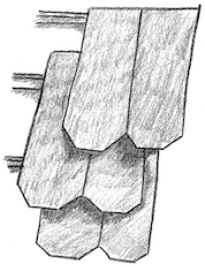
pick



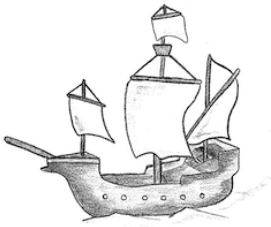
pig

prick

script



shingle



ship

shit

sing

sip

sit

skip

spit

spritz

sting

string

tip

trip

A.3 Consulted forums and articles for online meta-discourse of /s/-retraction

The online meta-discourse on /s/-retraction was conducted on March 30-31, 2019. A Google search was conducted for a modified spelling of the ten most frequent /str/ onset words, with ⟨sh⟩ in place of ⟨s⟩, as in ⟨shtreet⟩ instead of ⟨street⟩. The first ten pages of Google hits was examined for *street*, the most frequent /str/ word. The first five pages of hits were examined for the second through tenth most frequent /str/ words: *straight*, *strong*, *strange*, *destroy*, *strike*, *district*, *struck*, *strip*, *instructions*. Do to the relative prominence of *struggle* in the meta-discourse on line, five pages of hits were also examined for *shtruggle*.

The online forums and discussions that yielded insightful social commentary provided in this appendix, with the relevant comments reproduced below, unedited.

1. `https://www.antimoon.com/forum/t10923.htm`

    - My first judgement when I began to hear this peculiar sound was that young men didn't want to sound silibant or effeminate and so were trying to eliminate the hissing sound of "s".

2. `https://blogs.chicagotribune.com/news_columnists_ezorn/2011/10/shtay-shtrong-shtraight-talkers-youre-not-so-shtrange.html`

    - Those "shtr" people must be German.
    - Over the years, I've noticed that many of the groshery pronouncers are from around Cleveland.
    - As someone with an intermediate form (not *quite* to "shtr" but the 's' before a "tr" is not the same as the 's' elsewhere) I'd like to point out that the title of the post is a little off—it'd be "stay Shtrong".

3. `https://boards.straightdope.com/sdmb/showthread.php?t=548320`

    - In my case it's more like "schreet." The "ch" replaces the "t," not the "s."
    - Depends on how much I've had to drink.
    - I always thought it was a "ghetto" thing, like "axe" for "ask" or "A-ight" for "alright" I've noticed some people also say "sh-tupid". Co-incidence?
    - Could it be related to Yiddish in a way? I've noticed this occasionally from older Jews thhat are the children of immigrants; "schtreet" for street, "schtick" for stick (not used in reference to a comedy act or performance, but rather something like "schtick of butter"), and so on.
    - That's what it's been in my experience, too. I've lived in many parts of the country and have never heard it come from someone who wasn't 1) black and 2) from a ghetto of sorts.

- I wouldn't necessarily discount it, but I don't think so. The palatalization I hear and use only occurs in "str" clusters, not simple "st" clusters. So while I do say "shtring" and "shtreet," I don't say "shtop" and "shtick" or anything like that. Growing up in a very Polish area of Chicago, I don't think there was much Yiddish influence here in the accent, but it could certainly be one of those things that persisted or spread from other neighborhoods. The interesting thing is when I talk in a foreign language, I do not palatalize "str" clusters and they come out clear. It's only when I speak English (or German, as I was taught to do so in German.) And I'm surprised some people seem to have only heard it as a ghetto or black accent. It's pretty damned common among whites, too (see the list I provided in a previous post.)
- Folks from Hawaii who no longer speak pidgin or don't speak it in official situations usually still pronounce street or stream as shtreet or shtream.
- It is common in India for someone who did not grow up speaking English to pronounce it as "shtreet". Those who grew up speaking English pronounce it as "street".
- White female mid-20s middle-class Milwaukeean, and I say schreet. (Also, as people have observed above, schring, chree, and akchual. And drink is jrink.) Perfectly normal dialectical variation.
- I do the 'shtr' pronunciation, and I literally cannot say 'str'. I'm white, grew up in affluent suburbs of central NJ.
- Out here in California, I've noticed it's a kind of urban and ethnic thing. Aside from East Coast visitors, I hear it more from Mexicans or African Americans. But, when I lived in Hawai'i, it was common to hear "shtreet" and things like "frushtrated" nearly across the board. The exception would be malihini and haole (visitors and white folks).

4. `http://www.city-data.com/forum/general-u-s/1343085-replacing-d-t-accent-question-2.html`

    - Another African-American pronunciation that's creeping into general usage among all young people is using an "sh" sound when saying words that start with an "st" such as street ("sh-treet"), strong (sh-trong") and straight (sh-traight). I just started noticing it in the last 10-20 years - probably from the huge popularity of Hip-Hop and Rap culture spreading to the mainstream.

5. `https://dict.leo.org/forum/viewGeneraldiscussion.php?idforum=4&idThread=1156731&lp=ende&lang=en`

    - Northeast could be right; it's certainly not familiar to me down here in the south central US. Wonder if it has anything to do with German- or Yiddish-speaking populations.

- What are you listening to this audio book on? It is common for recording/playback equipment and some file-compression software to have troubles with sibilants in speech. Basically the high frequency segments of /s/ sounds are culled, leaving something that can sound like an /ʃ/.

6. https://english.stackexchange.com/questions/84836/why-is-str-sometimes-pronounced-as-shtr

   - As an example, Jay Z is rapping about the "New York city shtreets" (here at 0:30), not "streets". The actors from the Jersey Shore seem to do it as well, so I'm starting to believe this is an East Coast / New York thing?

7. https://www.englishforums.com/English/SoundBecomingFrequent/bbvwmh/post.htm

   - I have not noticed this trend among my fellow English speakers.
   - Neither have I, but don't rule out a physical impairment. People that wear dentures often complain that their "s" has changed to "sh."

8. https://www.facebook.com/6abcActionNews/posts/10156300507571378

   - Shtraight, shtreet, shtrong, shtruggle... That is not a speech impediment, it's a verbal crutch. You are professional broadcasters. Try harder to sound like ones.

9. https://forum.wordreference.com/threads/pronunciation-s-as-sh-in-ae-strong-street.3349230/

   - In American English, S is never pronounced as SH. That change doesn't happen in any dialect of American English.
   - It would be strange to deliberately adopt the "shtr-" pronunciation. Don't do that.

10. https://forum.wordreference.com/threads/pronunciation-of-str-%E2%86%92-shtr-medial.2575860/

    - I've just heard "indushtry" on the radio (no speech impediment). (There has been discussion for some time about the spread of initial str- being pronounced as shtr-: "shtrong, shtring, shtrangle" etc. This is the first time I've noticed it medially. Presumably this is a natural step in the spread?)
    - I'm pretty sure I've heard this on and off over the years - i.e. it doesn't strike me as either new or unusual (the s in unusual often takes in an h as it's said, but also sometimes loses its voicing by "picky pronouncers"!)). The difference between "a stream" and "industry" in terms of "medial" seems minor so, to me, it's hardly more remarkable than the initial str- picking up the h. There's a recent long thread on the h intruding in issue and it has some parallels with this question. The first parallel is simply that we acknowledge that it happens but we should not

encourage English learners to copy it. The second parallel is that the "amount" of h that intrudes, in my experience, is quite variable, from none to "hardly noticeable" to obvious - in the case of str- the sound becomes positively like standard German st- with a full-blown sh. In English, we don't have the cluster sr-vowel (as in Sri Lanka, so we insert the h to manage) For some, perhaps, it's the same with str-vowel.

11. https://www.grammarphobia.com/blog/2008/05/shtreet-smarts.html

    - This sound bugs me, but no one else seems to notice. Am I hearing things? Is this a regional or cultural dialect?

12. https://www.letsrun.com/forum/flat_read.php?thread=6130511

    - Curious speech pattern I've noticed in recent years - I'm pretty sure it started with African Americans, but now I hear almost everyone doing it (including those I've challenged doing so, but who have flatly denied it):
    - The OP is a doofus and made up this whole "shtr" business
    - If I asked a dozen people on the train where to get off for Greenwich Village, I'm reasonably confident that not a single one of them would tell me to get off at "4th Shtreet."
    - Every black person and most Jews and northeasterns says this, too. Listen if you don't believe me. I hate that I can hear this.
    - Why are you pretty sure it started with African Americans? Please explain this or is this just some racist garbage? I'm surprised so few people have attributed this to the Irish.
    - My experience is definitely different; I mostly hear this in white Americans with regional accents in the Northeast.
    - The "sh" / "s" substitution has been driving me crazy for some time now. "shtroller", "shtraight", "shtreet", "shtrong"...I can't believe that people aren't discussing this more frequently, even parodying it. It sends shivers up my spine. It seems to be especially prevalent on the East Coast, especially the New York area. Maybe it's just because I personally do not like that sound exchange, but I generally make an association with a lower socioeconomic and educational background. As much as I really, sincerely admire and love Michelle Obama, she does this all the time when she speaks. She has a few other rookie ticks—lots of slow and low "ahhh" sounds during pauses that are fairly distracting.

13. https://literalminded.wordpress.com/2011/09/06/shtraight-talk/

    - When my wife and sons and I were watching the movie Independence Day (1996), I heard Harry Connick Jr.'s character say to Will Smith's character, "You'll never get a chance to fly the space shuttle if you marry a shtripper." I made everyone wait while I rewound twice to make sure I'd heard right

- A month later, we were watching Beverly Hills Cop (1984), and I heard Eddie Murphy's character utter this other sentence about stripping: "The only reason these officers were in a shtrip club. . . ."
- A couple of weeks into the school year, I overheard a conversation among a couple of Adam's fellow fourth graders as they picked up their "Grab n Go" breakfast in the school hallway on the way to their classroom. Apparently the school can't count on parents actually giving their kids breakfast every morning, so they provide snacks before school for any kids who want them, so they can start off the day with something nutritious and be able to concentrate better in class. This morning, it was Pop Tarts. One girl said to another, "It was funny, because you said brown sugar and I said shtrawberry!" It really must have been funny, because the girl said it again, and again pronounced strawberry as shtrawberry.
- At about 7:51 into episode 414 of This American Life, the producer of the first story, Ben Calhoun, says, "These weren't regular uniformed cops. They were the guys in shtreet clothes." In the past year, I've heard one of each of Doug's and Adam's friends pronounce /str/ as [ʃtr], usually in the word destroy.
- During a family trip to New York City last month, a bus tour guide consistently pronounced /str/ as [ʃtr].
- I don't believe I use "shtr". If I recall rightly, I first noticed the phenomenon in one of my professors from the University of Calgary; he was from Saskatchewan.
- I never noticed this. I grew up in inner-city, working-class Columbus, OH, and I absolutely shtr. I have an Appalachian background as well, which is what people usually notice first, but perhaps I have a more linguistic mix than I first believed.
- I never noticed this until my wife pointed it out. I can never *not* hear it now. She has to repeat the word to me each time someone does it.
- I was just bitching about this weird trend in pronunciation on Facebook and someone linked me in here. Great stuff.The first time I ever remember hearing it was with Miley Cyrus on Hannah Montana. Just glad my daughter didn't pick it up.
- I first noticed this in some speakers of Black American English and George W. Bush, but when the announcer on public radio did it, I got upset. People look to public radio for correct standardized American pronunciations.
- When movie characters, such as those described above, use it, I think they are saying: Only pussies and mamma's boys would say straight when you can say shtraight. Straight is for fastidious librarians; shtraight is for guys who know how to load a gun and stuff a 20 down a shtripper's g-string. It's an anti-lisp. It says: Not only am I not gay, but I'm almost unbelievably shtraight.
- Does it mainly reveal a lack of education, say, from people on television, which young people then pick up and carry through to adulthood? Like others, I also find it nerve-wracking to hear but I think the reaction has to do with a fear that

inadvertently one might actually say it since it seems to happen almost naturally to those who

- I suspect this retroflex assimilation keeps going as long as there's a coronal available, regardless of morphology.
- I am not a linguist (only studied the subject freshman year in college), but on a very basic, layman's level, could the sound of "shtr" be appealing in some odd, absurd way? Like a lush, yummy (for lack of a better descriptor) sound that connotes something very pleasing? This is just a gut feeling and I'm lacking examples but maybe someone out there knows what I'm trying to describe...?
- It's also a local thing, right here in DC: The Mayor of Washington, DC, pronounces the name of the territory she's in charge of as the *Dishtrict of Columbia*. She's Muriel Bowser, and you can google her all over the place. Needless to say, she uses the word quite often....

14. `http://notmytribe.com/2008/shtrength-shtrong-shtreuth-82462.html`

    - Who says "SHtrong beside southern idiots like the current president? I now hear the god-awful mispronunciation on the lips of sub urban TV people, news reporters on location, even from the sports sidelines. Proper English is determined by usage, so nuclear will become nucular if enough yahoos say it's so.

15. `https://painintheenglish.com/case/5231`

    - Only the media could pick up and run with a complete misuse of pronunciation rules and thrust them into common usage. Even the army is doing this, with an obviously white announcer deliberately doing the 'shtrong' pronunciation in its recruitment ads targeted to black people.
    - I have been beefing about this to everyone I know since I first heard it from the first lady's campaign speech on the radio. I just heard it moments ago from the Charmin TP bear on tv commercial with Ultra Shtrong description. I thought it was just me but I found this article before I posted to Facebook. Someone has to make it stop!!!!!
    - I recall education Blacks in the 1960s using this pronunciation. It seems that when Blacks used this pronunciation back then it signaled an educated person. I just went back to recorded speeches of Shirley Chisholm. She used this pronunciation. Then from usage by education blacks that pronunciation seems to have entered into mainstream Black pronunciation and from there into common, widespread, mainshtream White usage. It's as fingernails on a blackboard for me. It's now 2016 and I'm noticing that the "h" is now being added to "st" strings as in "shtory". Anyone else follow this thread from educated Black speech? I know language changes. It's a natural process but my ear catches on every single pronunciation of "st" and "str" as "sht" and "shtr". The silliest pronunciation recently heard was largest trees pronounced as "largesht shtreesh". Totally mangled in other words and barely intelligible.

- This mispronunciationof the words like strong, and destroyed, by Michelle Obama has been so annoying and distracting and in my opinion really so unbecoming of a first lady. It also seems to me, that other words, like America, for example, are said with a tone of complaint or disdain. It is so distracting that I have trouble following the context of her remarks on a given occasion. As a role model for the youth of this nation, and speaking publicly as the First Lady, it surprises me no one ever counseled her on the inappropriateness of mispronounciation of these words that in my opinion, diminishes what she was trying to say in any given speech.
- Wow I can't believe someone besides me has noticed this ever-expanding trend! For years I've been pointing it out to others but constantly told it's in my imagination, that I'm just hearing it wrong.
- I have noticed the shtrong,etc. pronunciation for a while now and wondered if it was a physical difference in the tongue causing it. A new ad on TV features a black man saying "shprite" instead of Sprite.
- There are two answers to this concerning the white people who pronounciate certain words as so. 1) Some are deliberately 'mispronouncing' certain words that they know they can pronounce correctly. 2) Some actually have the tongue to pronounce these words as such beyond their control. I won't post the reasons why numbers 1 and 2 is possible (in the case of both, black people are included as well), you'll just have to think about it.
- I have noticed this phenomenon in Americans under 40 mostly actors . I have not noticed a solely African-American input except for when it is an African-American actor speaking the SHTR sound is much more distinguishable. Case in point , the lead actress in the new CBS Star Trek episode one is almost unwatchable . She even takes it beyond one word such as "this traitor" becomes "thish traitor."
- The first prominent person I noticed to speak this way was Michelle Obama, maybe it's still an Obama support thing with the media. Very shtrange.
- Sarah Huckabee Sanders almost exclusively pronounces "s" as "shtr". It's maddening and I can barely stand listening to her speak.
- I actually googled "pronunciation of strong" and came upon this site. It's something that I notice and wonder about often. I had a friend back in elementary school who pronounces "str" like "shtr." Every time I hear it, I think of her. I was guessing it's the way someone's mouth is shaped, like a minor speech impediment. I still don't know what to think of it but I hear it often.
- I'm glad to see that I wasn't imagining it. The link to the Penn research paper, shared by Cathy W. sheds some light on it. I believe its origin is with Polish immigrants in the early 20th Century, based in the Chicago area. The first time I ever heard this pronunciation is in the film "The Blues Brothers," which highlighted Chicago working-class speech patterns.

16. `https://pammarshalla.com/michelle-obamas-shtreet-for-street/`

- I have a 21-year-old [SLP] client with above average intelligence who says "shtreet" for "street." He also says "undershtanding" for "understanding" and "shtretch" for "stretch." He seems to do this on purpose. Any comments?
- In the 1970's, this substitution was heard primarily in "Southern Dialect" and it was standard in what was called "Black Dialect." Now this pattern seems to be generalizing to the broader population. With the president's wife using it, it would seem appropriate now to call it part of Standard North American English.

17. `http://people.sc.fsu.edu/~jburkardt/fun/wordplay/shtrange.html`

    - As long ago as the 1990's, when rap music spread across the musical horizon, I noticed a peculiarity of pronunciation in many recordings. Words beginning with "str" were clearly being pronounced as though they were spelled "shtr".
    - I suppose almost any minor language variation, given the right prominence, can be picked up unconsciously and repeated by others - after all, that's more or less how we learn to speak in the beginning, and it's how people renew and refresh their language, how Northerners who move south gradually pick up a partial Southern accent that seems to amuse both their old and new friends.

18. `https://www.quickanddirtytips.com/education/grammar/how-s-backing-causes-people-to-pronounce-street-as-schtreet`

    - IT IS NOT OKAY. You can offer explanations, but you cannot 'approve' of this. It is wrong.
    - Please stop making excuses for lazy pronunciation. It isn't a progressive change in the way society approaches a word with the 'st' phoneme. It's an abhorrent habit that is spreading because nobody will correct the error. As with the proclivity to pronounce the word 'strength' as 'strenth', this affectation is annoying, incorrect, and makes those that do it sound as though they have ill-fitting dentures or have had a bit too much to drink.
    - I have noticed it for the last couple years, and it's getting more widespread. It still sounds like a speech impediment to me, I think it is an impediment to children learning proper spelling and pronunciation. It feels more like slang, or almost an obsessive-compulsive disorder because I hear local anchors and weather people heavily using this S-backing, and it really impedes the point of content they are making. It's not correct- it's a foolish modern trend- and I wish it would go away.
    - This is a very interesting article with a compilation of takes on the subject. But, I have another theory: By the late 1800's, there were more Americans with German ancestry than those with English ancestry. In German, "str" and a few other such letter clusters is pronounced "shtr". I believe *that* is the main reason why words in the US are pronounced with s-backing.
    - How do you account for that fact that African-Americans have the highest incidence of this vocalization problem?

19. http://smith-wessonforum.com/lounge/527093-shtrolling-down-shtreet-shtraightaway.html

    - I can't stand it! It seems on the radio and TV, so many folks are turning a standard "s" into a a non-shtandard "sh." Do all these people live in mostly Germanic neighborhoods? Where did their get their English education?Stock becomes "shtock." Street becomes "shtreet." Strangle becomes "shtrangle." Stomach becomes "shtomach." Strange becomes "shtrange." It's driving me to dishtraction. It's my current pet peeve, and I never noticed it until just a few years ago - very prevalent now. My wife thought the first time she heard it that the person mispronouncing these words had a speech defect. But it's WAY more than just one person. It's got to SHTOP, and soon!
    - That's been a side effect of low quality on-location broadcasts, probably from cell phones, for a while now. Realized what it was when I saw a field report with the speech impediment followed by the same reporter from the studio, speaking clearly.
    - I heard this a lot from people in Pennsylvania, Ohio and upshtate New York..

20. https://spreemancommunications.com/tag/shtraight/

    - First I noticed radio DJs doing this. Then newscasters. Then teachers. But when Michelle Obama started talking about America's "shtruggles," everyone else picked it up shtraightaway

21. https://www.tek-tips.com/viewthread.cfm?qid=950657

    - This is something typically German. Perhaps a side effect after too much German beer?

22. https://www.urbandictionary.com/define.php?term=heightleaded

    - Wow, I feel sho heightleaded I cand even shtand up shtraight, led alone try ta walk a shtraight line!

23. https://www.waywordradio.org/discussion/topics/researchers-track-evolution-of-phillys-odd-accent/

    - I still listen to the Philadelphia local news on frequent occasion. One morning traffic reporter says "mash chransit" (mash chranzit) for "mass transit." The palatalization effect of the r keeps traveling back through the t all the way into the previous word and its final s. Even though I haven't lived in Philadelphia for nearly 40 years, when I hear mash chransit, I feel all warm inside.

24. https://www.waywordradio.org/street-vs-shtreet-and-straight-vs-shtraight/

- I don't hear it very often, but most recently there was a talk show host on a Chicago news station.

25. https://www.waywordradio.org/discussion/topics/street-vs-shtreet-why-and-where/

    - I have not noticed that at all. It seems to me like some actor affecting wearing denture.

26. https://www.waywordradio.org/discussion/topics/the-not-too-shtrange-shtr-pronunciation-of-str/

    - It is certainly common in Philadelphia and New York.

Separately, I have provided the meta-discourse from news articles, opinion, editorials, press releases, and radio shows, with the relevant comments pulled below.

1. https://www.americanthinker.com/articles/2012/09/obamas_struggle.html

    - On the campaign trail, Michelle Obama often employs the word "struggle." It was actually Mrs. Obama's peculiar pronunciation – "shtruggle" – that first drew my attention to the frequency of her usage.

2. https://www.chicagotribune.com/news/opinion/zorn/ct-minor-dishtress-20161125-column.html

    - Once you begin hearing people throwing extra h's into "str-" words, you can't stop hearing it.
    - It's not a sign of ignorance. Harvard-educated first lady Michelle Obama is "the worst violator of all" according to the Washington Post's Gene Weingarten.
    - It's not easier, faster, or smoother to inject a gratuitous "h" into "str" words. It's not a hoity-toity affectation or social signifier. It doesn't make the speaker sound extra casual or friendly. Shtrange, huh?

3. https://www.chronicle.com/blogs/linguafranca/2018/08/13/shtraight-talk-on-s-backing/

    - My late Aunt Eva, a white midwesterner, exhibited this speech pattern. I always presumed it was her loose dentures.
    - I first heard s-backing when I moved to Philadelphia from New England. I wondered whether the mistaken insertion of an "h" was a carryover from Pennsylvania Dutch, as German ancestry figures so heavily here and westward. (In German, "spiegel" is pronounced "shpeagle.") To this day, when I hear an s-backer, I assume the person is careless, unintelligent, or both.

4. https://news.chass.ncsu.edu/2016/04/21/research-examines-shtriking-sound-change-in-raleigh/

5. https://www.nytimes.com/1985/07/21/nyregion/stronger-urban-accents-in-northeast-are-called-sign-of-evolving-language.html

    - The new accents are so strong that it is possible, for example, to hear a sentence like this in Philadelphia... "He left his HAY-us in Northeast Ful-UFF-yuh and got into his core to GEH-oh DAY-un to Shpring GOR-den Shtreet like he oys does. But on the way he got into a FUH-eet with another driver. It was BEE-ad." Translation: "He left his house in Northeast Philadelphia and got into his car to go down to Spring Garden Street like he always does. But on the way he got into a fight with another driver. It was bad."

6. https://www.rushlimbaugh.com/daily/2016/12/20/michelle_obama_and_the_shtruggle/

    - Now, she's a leftist, and I know what the shtruggle is. And frankly, folks, I'm growing weary of it. "The shtruggle" is the premise that minorities do not have a prayer in America. The deck is so stacked against them that they don't have a chance. And this shtruggle to overcome this great injustice is never ending. She speaks of the shtruggle — and you have to call it the shtruggle, not the struggle. It's s-h, the shtruggle. Every time she makes a speech. It is a feature of practically every speech that she makes, and what does it do? It reinforces the legitimacy of the shtruggle.

7. https://www.washingtonpost.com/lifestyle/magazine/gene-weingarten-admits-his-personal-shtruggle/2016/04/27/8ae99776-fcda-11e5-9140-e61d062438bb_story.html

    - My point is, no English word should be pronounced with a "shtr" sound. Yet suddenly, around 2001 or thereabouts, I began hearing politicians talk about the need for "shtrength against terrorism." We launched "air shtrikes." This linguistic hiccup was particularly rancid to my ears. For one thing, it sounds shtupid. But there is also a faint echo of . . . Hitler.
    - I blame the orthodontists, who have been malforming youngshters' mouths for years (probably just during the period when this shtriking change appeared). Some god or other gave us crooked teeth so that we would shtruggle with pronunciation in our formative years and learn correctness (not political), but these mouth-revisors have hammered and twisted our shpeaking tools with the goal of making us all the same when we shmile, right down to the shparkle.

A.4 Cognitive, personality, & demographic surveys

In this section of the appendix, I have reproduced the post-test demographic survey completed by the participants. If no options are not provided (in italics), questions are assumed to be free response unless otherwise specified.

A.4.1 Basic demographic questions

1. In what year were you born?
2. How would you describe your gender identity? *Female, male, non-binary, other (please specify)*
3. What was your sex assigned at birth? *Female, male, intersex or other*
4. How would you describe you sexual orientation? *Straight or heterosexual, gay, lesbian or homosexual, bisexual, queer, asexual, other (please specify)*
5. How would you describe your race? (Select all that apply) *African American/Black, Asian, Hispanic/Latino/a, Middle Eastern/Arab, Native American/American Indian, Native Hawaiian/Pacific Islander, South Asian/Indian, White/European/Caucasian, other (please specify)*
6. Did you grow up in a household where all members spoke primarily English until the time you were 12? *Yes, no, not sure*
7. If No, what language and how frequently was English spoken?
8. Do you speak or have you studied any languages other than English? *Yes, no, not sure*
9. If Yes, what language(s) and for how long?
10. Would you describe yourself as a native speaker of North American English? *Yes, no, not sure*
11. In what state were you born?
12. If you were born outside the US/Canada, where were you born?
13. In what state/province have you lived the longest before age 18?
14. If you were raised outside the US/Canada, where did you live the longest before age 18?
15. In what city/town did you live the longest during that time?
16. How would your describe the environment where you lived during that time? *Urban, suburban, rural*
17. Please list any other states, provinces or countries you have lived in for greater than a year.

A.4.2 Neurological, language, and hearing questions

18. Do you have any difficulty with the following? (select all that apply) *Talking, finding words, understanding speech, hearing, reading, writing, none of the above*

19. As far as you know, do you have any neurological impairments? If so, please specify, e.g. Parkinson's, Huntington's, stroke, epilepsy, etc.

20. As far as you know, do you have normal hearing? *Yes, no, not sure*

21. Do you wear contacts or glasses to have corrected-to-normal vision? *I do NOT require corrective lenses, my vision is normal; I wear SINGLE (i.e. not bifocal) corrective lenses, my vision is corrected-to-normal; I wear BIFOCAL corrective lenses, my vision is corrected-to-normal; I require corrective lenses, but I'm NOT wearing them today; I have a visual impairment, my vision cannot be corrected-to-normal*

22. Are you left-handed or right-handed? *Left-handed, right-handed, comfortable with both*

A.4.3 Big Five

The forty-three question Big Five Inventory (John et al., 2008) was included to assess personality traits along five dimensions: Extroversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. The following statements were provided with radio buttons *Strongly agree*, *Slightly agree*, *Slightly disagree*, and *Strongly disagree*. Participants were asked: I see myself as someone who...

23. Is talkative.

24. Tends to find fault with others.

25. Does a thorough job.

26. Is depressed/blue.

27. Is original/comes up with new ideas.

28. Is reserved.

29. Is helpful and unselfish with others.

30. Can be somewhat careless.

31. Is relaxed/handles stress well.

32. Is curious about many different things.

33. Is full of energy.

34. Starts quarrels with others.

35. Is a reliable worker.
36. Can be tense.
37. Is ingenious/a deep thinker.
38. Generates a lot of enthusiasm.
39. Has a forgiving nature.
40. Tends to be disorganized.
41. Worries a lot.
42. Has an active imagination.
43. Tends to be quiet.
44. Is generally trusting.
45. Tends to be lazy.
46. Is emotionally stable/not easily upset.
47. Is inventive.
48. Has an assertive personality.
49. Can be cold and aloof.
50. Perseveres until the task is finished.
51. Can be moody.
52. Values artistic, aesthetic experiences.
53. Is sometimes shy or inhibited.
54. Is considerate and kind to almost everyone.
55. Does things efficiently.
56. Remains calm in tense situations.
57. Prefers work that is routine.
58. Is outgoing and sociable.
59. Is sometimes rude to others.
60. Makes plans and follows through with them.

61. Gets nervous easily.
62. Likes to reflect/play with ideas.
63. Has few artistic interests.
64. Likes to cooperate with others.
65. Is easily distracted.
66. Is sophisticated in art, music or literature

A.4.4 Empathy Quotient

The twenty-two question Empathy Quotient survey (EQ: Baron-Cohen & Wheelwright, 2004) was included to assess the degree to which respondents are able to identify and respond to another individual's thoughts and emotions. The following statements were provided with radio buttons *Strongly agree*, *Slightly agree*, *Slightly disagree*, and *Strongly disagree*. Participants were asked: Please respond to the following statements as honestly as possible.

67. I can easily tell if someone else wants to enter a conversation.
68. I really enjoy caring for other people.
69. I find it hard to know what to do in a social situation.
70. I often find it difficult to judge if something is rude or polite.
71. In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.
72. I can pick up quickly if someone says one thing but means another.
73. It is hard for me to see why some things upset people so much.
74. I find it easy to put myself in somebody else's shoes.
75. I am good at predicting how someone will feel.
76. I am quick to spot when someone in a group is feeling awkward or uncomfortable.
77. I can't always see why someone should have felt offended by a remark.
78. I don't tend to find social situations confusing.
79. Other people tell me I am good at understanding how they are feeling and what they are thinking.
80. I can easily tell if someone else is interested or bored with what I am saying.

81. Friends usually talk to me about their problems as they say I am very understanding.
82. I can sense if I am intruding, even if the other person doesn't tell me.
83. Other people often say that I am insensitive, though I don't always see why.
84. I can tune into how someone else feels rapidly and intuitively.
85. I can easily work out what another person might want to talk about.
86. I can tell if someone is masking their true emotion.
87. I am good at predicting what someone will do.
88. I tend to get emotionally involved with a friend's problems.

A.4.5 MRAS

The ten question MRAS survey (Pleck et al., 1993) was included to measure participant's relative endorsement of traditional stereotypes of masculinity. The following statements were provided with radio buttons *Strongly agree*, *Slightly agree*, *Slightly disagree*, and *Strongly disagree*. Participants were asked: Please respond to the following statements as honestly as possible.

89. It is essential for a guy to get respect from others.
90. A man always deserves the respect of his wife and children.
91. I admire a guy who is totally sure of himself.
92. A guy will lose respect if he talks about his problems.
93. A young man should be physically tough, even if he's not big.
94. It bothers me when a guy acts like a girl.
95. I don't think a husband should have to do housework.
96. Men are always ready for sex.
97. The thought of men having sex with each other is disgusting.
98. I could never be friends with a gay man.

A.4.6 Promis Anxiety

The twenty-nine question PROMIS Anxiety score (Cella et al., 2010) was included to assess self-reported anxiety. The following statements were provided with radio buttons *Never*, *Rarely*, *Sometimes*, *Often* and *Always*. Participants were asked: Please respond to the following statements as honestly as possible. In the past 7 days...

99. I felt fearful.
100. I felt frightened.
101. It scared me when I felt nervous.
102. I felt anxious.
103. I felt like I needed help for my anxiety.
104. I was concerned about my mental health.
105. I felt upset.
106. I had a racing or pounding heart.
107. I was anxious if my normal routine was disturbed.
108. I had sudden feelings of panic.
109. I was easily startled.
110. I had trouble paying attention.
111. I avoided public places or activities.
112. I felt fidgety.
113. I felt something awful would happen.
114. I felt worried.
115. I felt terrified.
116. I worried about other people‘s reactions to me.
117. I found it hard to focus on anything other than my anxiety.
118. My worries overwhelmed me.
119. I hard twitching or trembling muscles.
120. I felt nervous.

121. I felt indecisive.

122. Many situations made me worry.

123. I had difficulty sleeping.

124. I had trouble relaxing.

125. I felt uneasy.

126. I felt tense.

127. I had difficulty calming down.

A.4.7 Experimental Impressions

128. Do you have any guesses at what the experiment was looking at?

The following statements were provided with radio buttons *Strongly agree*, *Slightly agree*, *Slightly disagree*, and *Strongly disagree*. Participants were asked: What is your impression of the voice that gave you instructions during the image identification task?

129. He sounded friendly.

130. He sounded masculine.

131. He sounded outgoing.

132. He sounded casual.

133. He sounded gay.

134. He sounded attractive.

135. He sounded educated.

REFERENCES

Allopenna, Paul D, James S. Magnuson & Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38(4). 419–439. https://doi.org/10.1006/jmla.1997.2558.

Altendorf, Ulrike. 2003. *Estuary English: Levelling at the interface of RP and South-Eastern British English*. Tübingen: Gunter Narr.

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson & Regina Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech* 34(4). 351–366. groups.inf.ed.ac.uk/maptask/.

Arbib, Michael. 2002. The mirror system, imitation, and the evolution of language. In Chrystopher Nehaniv & Kerstin Dautenhahn (eds.), *Language origins: Perspectives on evolution*, 229–280. MIT Press.

Arbib, Michael. 2005. The mirror system hypothesis: How did protolanguage evolve? In M Tallerman (ed.), *Language origins: Perspectives on evolution*, 21–47. Oxford University Press.

Archangeli, Diana, Adam Baker & Jeff Mielke. 2011. Categorization and features evidence from American English /ɹ/. In G N Clements & R Ridouane (eds.), *Where do phonological features come from?: Cognitive, physical and developmental bases of distinctive speech categories*, 173–196. Amsterdam: John Benjamins. https://doi.org/10.1075/lfab.6.

Auer, Peter & Frans Hinsken. 2005. The role of interpersonal accommodation in a theory of language change. In P. Auer, F. Hinsken & P. Kerswill (eds.), *Dialect change: convergence and divergence in European languages*, 335–357. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486623.

Auer, Peter & Frans Hinskens. 1996. The convergence and divergence of dialects in europe. new and not so new developments in an old area. *Sociolinguistica* 10(1). 1–30. https://doi.org/10.1515/9783110245158.1.

Babel, Molly. 2009. *Phonetic and social selectivity in speech accommodation*: University of California, Berkeley dissertation. https://escholarship.org/uc/item/1mb4n1mv.

Babel, Molly. 2010. Dialect divergence and convergence in new zealand english. *Language in Society* 39(04). 437–456. https://doi.org/10.1017/S0047404510000400.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40. 177–189. https://doi.org/10.1016/j.wocn.2011.09.001.

Babel, Molly & Dasha Bulatov. 2011. The role of fundamental frequency in phonetic accommodation. *Language and Speech* 55(2). 231–248. https://doi.org/10.1177/0023830911417695.

Baker, Adam, Diana Archangeli & Jeff Mielke. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23. 347–374. https://doi.org/10.1017/S0954394511000135.

Bargh, John A. & Erin L. Williams. 2006. The automaticity of social life. *Current Directions in Psychological Science* 15(1). 1–4. https://doi.org/10.1111/j.0963-7214.2006.00395.x.

Baron-Cohen, Simon & Sally Wheelwright. 2004. The Empathy Quotient: An investigation of adults with Asperger Syndrome or High Functioning Autism and normal sex differences. *Journal of Autism and Developmental Disorders* 34. 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00.

Bass, Michael. 2009. Street or shtreet? Investigating (str-) patalisation in Colchester English. *Estro: Essex Student Research Online* 1. 10–21.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Stastical Software* 67(1). 1–48. https://arxiv.org/abs/1406.5823.

Bauer, Laurie & Paul Warren. 2004. New Zealand English: Phonology. In B. Kortmann, E. W. Schnieder, K. Burridge, R. Mesthrie & C. Upton (eds.), *A handbook of varities of English, a multimedia reference tool*, vol. 1, 580–602. Berlin: Mouton de Gruyter.

Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85(4). 785–821. https://doi.org/10.1353/lan.0.0165.

Beddor, Patrice Speeter, James Harnsberge & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30(4). 591–627. https://doi.org/10.1006/jpho.2002.0177.

Beddor, Patrice Speeter & Rena Arens Krakow. 1999. Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America* 106(5). 2868–2887. https://doi.org/10.1121/1.428111.

Beddor, Patrice Speeter, Kevin B. McGowan, Julie E. Boland, Andries W. Coetzee & Anthony Brasher. 2013. The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133(4). 2350–2366. https://doi.org/10.1121/1.4794366.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3). 355–387. http://doi.org/10.1016/0010-0285(86)90004-6.

van Borsel, John, Els De Bruyn, Evelien Lefebvre, Anouschka Sokoloff, Sophia De Ley & Nele Daudonck. 2009. The prevalence of lisping in gay men. *Journal of Communication Disorders* 42. 100–106. https://doi.org/10.1016/j.jcomdis.2008.08.004.

Branigan, Holly P., Martin J. Pickering & Alexandra A. Cleland. 2000. Syntactic coordination in dialogue. *Cognition* 75(2). B13 – B25. https://doi.org/10.1016/S0010-0277(99)00081-5.

Brennan, Susan E. & Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22(6). 1482–1493. http://doi.org/10.1037/0278-7393.22.6.1482.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977–990. https://doi.org/10.3758/BRM.41.4.977.

Bukmaier, Véronique, Jonathan Harrington & Felicitas Kleber. 2014. An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: evidence for a gradient sound change. *Frontiers in Psychology* 5. 5–12. https://doi.org/10.3389/fpsyg.2014.00828.

Campbell-Kibler, Kathryn. 2011a. Intersecting variables and perceived sexual orientation in men. *American Speech* 86(1). 52–68. https://doi.org/10.1215/00031283-1277510.

Campbell-Kibler, Kathryn. 2011b. The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22. 423–441. https://doi.org/10.1017/S0954394510000177.

Campbell-Kibler, Kathryn. forthcoming. The cognitive structure of indexicality: Correlations between tasks linking /s/ and masculinity. In L. Hall-Lew, E. Moore & R. J. Podesva (eds.), *Social meaning and linguistic variation: Theorizing the third wave*, .

Carnegie Mellon University. 2008. CMUdict: The CMU pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Cella, David, William Riley, Arthur Stone, Nan Rothrock, Bryce Reeved, Susan Yount, Dagmar Amtmann, Rita Bode, Daniel Buysse, Seung Choi, Karon Cook, Robert DeVellis, Darren DeWalth, James F. Fries, Richard Gershon, Elizabeth A. Hahn, Jin-Shei Lai, Paul Pilkonis, Dennis Revicki, Matthias Rose, Kevin Weinfurt, Ron Hays & PROMIS Cooperative Group. 2010. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology* 63(11). 1179–1194. https://doi.org/10.1016/j.jclinepi.2010.04.011.

Chambers, Jack. 1992. Dialect acquisition. *Langauge* 68. 673–705. https://doi.org/10.1353/lan.1992.0060.

Charman, Tony. 2006. Imitation and the development of language. In *Imitation and the social mind: Autism and typical development*, 96–17. New York: Guilford.

Chartrand, Tanya L. & John A Bargh. 1999. The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology* 76(6). 893–910. https://doi.org/10.1037/0022-3514.76.6.893.

Clarke, Sandra. 2004. A note on several unusual fricative pronunciations on the southwest coast of Newfoundland. *Regional Language Studies... Newfoundland* 18. 15–17.

Connine, Cynthia M. & Laura M. Darnieder. 2009. Perceptual learning of co-articulation in speech. *Journal of Memory and Language* 61(3). 368–378. https://doi.org/10.1016/j.jml.2009.07.003.

Crist, Sean. 1997. Duration of onset consonants in gay male stereotypes speech. In A. Dimitriadis, H. Lee, L. Siegel & A. Williams (eds.), *University of Pennsylvania Working Papers in Linguistics*, vol. 4 3, 53–70. https://repository.upenn.edu/pwpl/vol4/iss3/4/.

Cruttenden, Alan. 2014. *Gimson's pronunciation of English (8th edition)*. New York: Routledge.

Darwin, Chris. 2005. *Digital mixing script*. University of Sussex.

Delattre, Pierre & Donald C. Freeman. 1968. A dialect study of American r's X-ray motion picture. *Linguistics* 44. 29–68. https://doi.org/10.1515/ling.1968.6.44.29.

Delvaux, Véronqiue & Alain Soquet. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64. 145–173. https://doi.org/10.1159/000107914.

DiCanio, Christian. 2013. *Time averaging for fricatives*. Haskins Laboratories and SUNY Buffalo. http://www.acsu.buffalo.edu/$\sim$cdicanio/scripts.html.

Dijksterhuis, Ap & John A Bargh. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. Zanna (ed.), *Advances in experimental social psychology*, 1–40. San Diego, CA: Academic Press. https://doi.org/10.1016/S0065-2601(01)80003-4.

Dufour, Sophie & Noël Nguyen. 2013. How much imitation is there in a shadowing task? *Frontiers in Psychology* 4. 346. https://doi.org/10.3389/fpsyg.2013.00346.

Durian, David. 2007. Getting stronger every day?: More on urbanization and the socio-geographic diffusion of (str) in Columbus, OH. In S. Brody, M. Friesner, L. Mackenzie & J. Tauberer (eds.), *University of Pennsylvania Working Papers in Linguistics*, vol. 13 2, 65–79. https://repository.upenn.edu/pwpl/vol13/iss2/6/.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x.

Elman, Jeffrey & John McClelland. 1986. Exploiting the lawful variability in the speech wave. In J S Perkell & D Klatt (eds.), *Invariance and variability of speech processes*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Feldman, Naomi H., Thomas L. Griffiths & James L Morgan. 2009. The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological review* 116(4). 752–782. https://doi.org/10.1037/a0017196.

Fowler, Carol. 1996. Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99. 1730–1741. https://doi.org/10.1121/1.415237.

Fuchs, Susanne & Martine Toda. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors. In *Turbulent sounds: An interdisciplinary guide*, 281–302. Berlin: De Gruyter. https://doi.org/10.1515/9783110226584.281.

Fujisaki, H. & T. Kawashima. 1969. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute* 67–72.

Galle, Marcus E., Jamie Klein-Packard, Kayleen Schreiber & Bob McMurray. 2019. What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory. *Cognitive Science* 43. e12700. https://doi.org/10.1111/cogs.12700.

Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance* 6(1). 110–125. https://doi.org/10.1037/0096-1523.6.1.110.

Garrod, Simon & Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27(2). 181 – 218. https://doi.org/10.1016/0010-0277(87)90018-7.

Garrod, Simon & Martin J. Pickering. 2004. Why is conversation so easy? *Trends in Cognitive Sciences* 8(1). 8–11. https://doi.org/10.1016/j.tics.2003.10.016.

Gick, Bryan & Donald Derrick. 2009. Aero-tactile integration in speech perception. *Nature* 462. 502–504. https://doi.org/10.1038/nature08572.

Giles, Howard. 1973. Accent mobility: a model and some data. *Anthropological Linguistics* 15. 87–105. https://www.jstor.org/stable/30029508.

Giles, Howard, Nikolas Coupland & Justine Coupland. 1991. Accomodation theory: Communication, context and consequence. In H. Giles, J. Coupland & N. Coupland (eds.), *Contexts of accomodation*, 1–68. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511663673.

Giles, Howard & Philip Smith. 1979. Accommodation theory: Optimal levels of convergence. In H. Giles, S. Clair & N. Robert (eds.), *Languages and social psychology*, Basil Blackwell.

Giles, Howard, Donald Taylor & Richard Bourhis. 1973. Towards a theory of interpersonal accommodation through language: Some Canadian data. *Langauge in Society* 2(2). 177–192. https://doi.org/10.1017/S0047404500000701.

Gilles, Peter. 1998. Virtual convergence and dialect levelling in luxembourgish. *Folia Linguistica* 32(1-2). https://doi.org/10.1515/flin.1998.32.1-2.69.

Glain, Olivier. 2013. *Les cas de palatalisation contemporaine (CPC) dans le monde anglophone*: Université Jean Moulin (Lyon 3) dissertation. http://www.theses.fr/2013LYO30053.

Glain, Olivier. 2014. Introducing contemporary palatalisation. In T. Lee (ed.), *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York Conference*, 16–29.

Goldinger, Stephen D. 1997. Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (eds.), *Talker variability in speech processing*, 33–66. San Diego, CA: Academic Press.

Goldinger, Stephen D. 1998. Echoes of echoes? an episodic theory of lexical access. *Psychological review* 105(2). 251–279. https://doi.org/10.1037/0033-295x.105.2.251.

Goldinger, Stephen D & Tamiko Azuma. 2003. Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics* 31(3). 305–320. https://doi.org/10.1016/S0095-4470(03)00030-5.

Goldinger, Stephen D & Tamiko Azuma. 2004. Episodic memory in printed word naming. *Psychonomic Bulletin & Review* 11(4). 716–722. https://doi.org/10.3758/BF03196625.

Gratier, Maya & Emmanuel Devouche. 2011. Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Development Psychology* 47(1). 67–76. https://doi.org/10.1037/a0020722.

Gregory, Stanford W & Brian R Hoyt. 1982. Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research* 11(1). 35–46. https://doi.org/10.1007/BF01067500.

Gregory, Stanford W, Stephen Webster & Gang Huang. 1993. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication* 13(3). 195–217. https://doi.org/10.1016/0271-5309(93)90026-J.

Gylfadottir, Duna. 2015. Shtreets of Philadelphia: An acoustic study of /str/-retraction in a naturalistic speech corpus. In *University of Pennsylvania Working Papers in Linguistics*, vol. 21 2, 2–11. http://repository.upenn.edu/pwpl/vol21/iss2/11/.

Haley, Katarina L., Elizabeth Seelinger, Kerry Callahan Mandulak & David J. Zajac. 2010. Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics* 38(4). 548–554. https://doi.org/10.1016/j.wocn.2010.07.006.

Hall, T. A., Silke Hamann & Marzena Zygis. 2006. The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association* 36(1). 59–81. https://doi.org/10.1017/S0025100306002453.

Harrington, Jonathan. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34. 436–457. http://doi.org/10.1016/j.wocn.2005.08.001.

Harrington, Jonathan, Michele Gubian, Mary Stevens & Florian Schiel. 2019. Phonetic change in an antarctic winter. *Journal of the Acoustical Society of America* 146(5). 3327–3332. https://doi.org/10.1121/1.5130709.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: an acoustic and perceptual study. *Journal of the Acoustical Society of America* 123(5). 2825–2835. https://doi.org/10.1121/1.2897042.

Harrington, Jonathan, Sallyanne Palethrope & Catherine I. Watson. 2000a. Does the Queen speak the Queen's English? *Nature* 408(6815). 927–928. https://doi.org/10.1038/35050160.

Harrington, Jonathan, Sallyanne Palethrope & Catherine I. Watson. 2000b. Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association* 30(1-2). 63–78. https://doi.org/10.1017/S0025100300006666.

Hinrichs, Lars, Axel Bohmann, Erica Brozovsky, Noli Chew, Kirsten Meemann & Patrick Schultz. 2016. Sibilants and ethnic diversity: A sociophonetic study of paltalized /s/ in STR-clusters among Hispanic, White and African-American speakers of Texas English. Talk presented at the Sixteenth Annual Meeting of the Texas Linguistics Society, Austin.

Hinskens, Frans. 1996. *Dialect levelling in Limburg: Structural and sociolinguistics aspects*. Tübingen: Max Niemeyer Verlag.

Hombert, Jean-Marie. 1977. Development of tones from vowel height? *Journal of Phonetics* 5. 9–16. https://doi.org/10.1016/S0095-4470(19)31109-X.

Hughes, George W & Morris Halle. 1956. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America* 28. 303–310. https://doi.org/10.1121/1.1908271.

John, Oliver P., Laura P. Naumann & Christopher J. Soto. 2008. Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P.

John, R. W. Robins & L. A. Pervin (eds.), *Handbook of personality: Theory and research (3rd edition)*, 114–158. New York: Guilford. https://doi.org/10.1121/1.1288413.

Jongman, Allard, Ratree Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108(3). 1252–1263. https://doi.org/10.1121/1.1288413.

Kataoka, Reiko. 2009. A study on perceptual compensation for /u/-fronting in American English. In *UC Berkeley Phonology Lab Annual Report*, 210–223. Berkeley: University of California–Berkeley.

Kim, Midam, William S. Horton & Ann R. Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2(1). 125–156. https://doi.org/10.1515/labphon.2011.004.

Kingston, John, Joshua Levy, Amanda Rysling & Adrian Staub. 2016. Eye movement evidence for an immediate ganong effect. *Journal of Experimental Psychology: Human Perception & Performance.* 42(12). 1969–1988. http://doi.org/10.1037/xhp0000269.

Klatt, Dennis. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language and Hearing Research* 18(4). 686–706. https://doi.org/10.1044/jshr.1804.686.

Kleber, Felicitas, Jonathan Harrington & Ulrich Reubold. 2011. The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech* 55(3). 383–405. https://doi.org/10.1177/0023830911422194.

Kraljic, Tanya, Susan E. Brennan & Arthur G. Samuel. 2008. Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107(1). 54–81. https://doi.org/10.1016/j.cognition.2007.07.013.

Kuhl, Patricia K & Andrew N Meltzoff. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America* 100(4). 2425–2438. https://doi.org/10.1121/1.417951.

Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer typologie des lautwandels und ihre konsequenzen.* Wiesbaden: Reichert.

Labov, William. 1984. Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–53. Englewood Cliffs, NJ: Prentice-Hall.

Labov, William. 2001. *Principles of Linguistic Change, Vol. II: Social Factors.* Malden, MA: Blackwell.

Lawrence, Wayne P. 2000. /str/ → /ʃtr/: Assimilaton at a distance? *American Speech* 75(1). 82–87. https://doi.org/10.1215/00031283-75-1-82.

Levon, Erez. 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society* 36. 533–554. https://doi.org/10.1017/S004740450707043.

Levon, Erez. 2014. Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society* 43(5). 539–566. https://doi.org/10.1017/S0047404514000554.

Lewandowski, Natalie & Matthias Jilka. 2019. Phonetic convergence, language talent, personality and attention. *Frontiers in Communication* 4. https://doi.org/10.3389/fcomm.2019.00018.

Liberman, Alvin M, Katherine Safford Harris, Howard S. Hoffman & Belver C Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54(5). 358–368. https://doi.org/10.1037/h0044417.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-2037-8_16.

Lindblom, Björn, Augstine Aguwele, Harvey Sussman & Elisabet Eir Cortes. 2007. The effect of emphatic stress on consonant vowel coarticulation. *Journal of the Acoustical Society of America* 121(6). 3802–3813. https://doi.org/10.1121/1.2730622.

Linville, Sue Ellen. 1998. Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopædica* 50. 35–48. https://doi.org/10.1159/000021447.

Ma, Debbie S, Joshua Correll & Bernd Wittenbrink. 2015. The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods* 47(4). 1122–1135. https://doi.org/10.3758/s13428-014-0532-5.

Mack, Sara & Benjamin Munson. 2012. The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics* 40. 198–212. https://doi.org/10.1016/j.wocn.2011.10.002.

Mann, Virginia & Bruno H Repp. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28(3). 212–228. https://doi.org/10.3758/BF03204377.

Manuel, Sharon. 1990. The role of contrast in limit vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88. 1286–1298. https://doi.org/10.1121/1.399705.

Martin, James G. & H. Timothy Bunnell. 1981. Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America* 69(2). 559–567. https://doi.org/10.1121/1.2018008.

Masur, Elise Frank & Doreen L. Eichorst. 2002. Infants' spontaneous imitation of novel versus familiar words: Relations to observational and maternal report measures of their lexicons. *Merrill-Palmer Quarterly* 48(4). 405–426. https://doi.org/10.1353/mpq.2002.0019.

Matthies, Melanie L., Pascal Perrier, Joseph S. Perkell & Majid Zandipour. 2001. Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language and Hearing Research* 44. 340–353. https://doi.org/10.1044/1092-4388(2001/028).

Mattys, Sven L., F. Seymour, Angela S. Attwood & Marcus R. Munafò. 2013. Effects of acute anxiety induction on speech perception: Are anxious listeners distracted listeners? *Psychological Science* 24. 1606–1608. https://doi.org/10.1177/0956797612474323.

McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748. https://doi.org/10.1038/264746a0.

McMurray, Bob, Meghan Clayards, Michael K. Tanenhaus & Richard N. Aslin. 2008. Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review* 15(6). 1064–1071. https://doi.org/10.3758/PBR.15.6.1064.

McMurray, Bob & Allard Jongman. 2011. What information is necessary for speech categorization? harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological review* 118(2). 219–246. https://doi.org/10.1037/a0022325.

Meltzoff, Andrew N & M. Keith Moore. 1977. Imitation of facial and manual gestures by human neonates. *Science* 198(4312). 75–78. https://doi.org/10.1126/science.198.4312.75.

Mielke, Jeff, Adam Baker & Diana Archangeli. 2010. Variability and homogeneity in American English /r/ allophony and /s/ retraction. *Laboratory Phonology* 10. 699–719.

Mielke, Jeff, Adam Baker & Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92(1). 101–140. https://doi.org/10.1353/lan.2016.0019.

Mitterer, Holger & Mirjam Ernestus. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109(1). 168 – 173. https://doi.org/10.1016/j.cognition.2008.08.002.

Mitterer, Holger & Jochen Müsseler. 2013. Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception, & Psychophysics* 75(3). 557–575. https://doi.org/10.3758/s13414-012-0407-8.

Munson, Benjamin. 2010. Variation, implied pathology, social meaning, and the 'gay lisp': A response to van Borsel et al. (2009). *Journal of Communication Disorders* 43. 1–5. https://doi.org/10.1016/j.jcomdis.2009.07.002.

Namy, Laura L, Lynne C Nygaard & Denise Sauerteig. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21(4). 422–432. https://doi.org/10.1177/026192702237958.

Natale, Michael. 1975a. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32(5). 790–804. https://doi.org/10.1121/1.2178720.

Natale, Michael. 1975b. Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills* 827–830. https://doi.org/10.2466/pms.1975.40.3.827.

Nguyen, Noël & Véronqiue Delvaux. 2015. Role of imitation in the emergence of phonological systems. *Journal of Phonetics* 53. 46–54. https://doi.org/10.1016/j.wocn.2015.08.004.

Nieldielzki, Nancy & Howard Giles. 1996. Linguistic accomodation. In H. Goebl, P. H. Nelde, Z. Stary & W. Wölck (eds.), *Kontaktlinguistik: Ein internationales Handbuch zeitgenössischer Forschung [an international handbook of contemporary research]*, vol. 1, 332–342. Berlin: Mouton de Gruyter.

Nielsen, Kuniko. 2008. *Word-level and feature-level effects in phonetic imitation*: University of California, Los Angeles dissertation.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39. 132–142. https://doi.org/10.1016/j.wocn.2010.12.007.

Nittrouer, Susan. 1995. Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America* 97. 520–530. https://doi.org/10.1121/1.412278.

Ohala, John. 1993. The phonetics of sound change. In C. Jones (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman.

Ohala, John & Maria-Josep Solé. 2010. Tubulence & phonology. In Susanne Fuchs, Martine Toda & Marzena Zygis (eds.), *Turbulent sounds: An interdisciplinary guide,*, 37–97. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110226584.37.

Olive, Joseph P., Alice Greenwood & John Coleman. 1993. *Acoustics of American English speech: A dynamic approach*. New York: Springer.

Ostreicher, Harvey & Donald Sharf. 1976. Effects of coarticulation on the identification of deleted consonant and vowel sounds. *Journal of Phonetics* 4. 285–301. https://doi.org/10.1016/S0095-4470(19)31256-2.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119. 2382–2393. https://doi.org/10.1121/1.2178720.

Pardo, Jennifer S. 2009. Expressing oneself in conversational interaction. In *Expressing oneself/expressing one's self: Communication, cognition, language, and identity*, 183–196. London: Taylor & Francis.

Pardo, Jennifer S., Rachel Gibbons, Alexandra Suppes & Robert M. Krauss. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40(1). 190–197. https://doi.org/10.1016/j.wocn.2011.10.001.

Pardo, Jennifer S, Isabel Cajori Jay & Robert M. Krauss. 2010. Conversational role influences speech imitation. *Attention, Perception & Psychophysics* 72(8). 2254–2264. https://doi.org/10.3758/APP.72.8.2254.

Pardo, Jennifer S, Kelly Jordan, Rollienne Mallari, Caitlin Scanlon & Eva Lewandowski. 2013. Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language* 69(3). 183–195. https://doi.org/10.1016/j.jml.2013.06.002.

Paul, Hermann. 1870, 5th edn. 1920. *Prinzipien der sprachgeschichte*. Halle: Niemeyer 5th edn.

Paul, Hermann. 1888. *Principles of the history of language*. London: Swan Sonnenschein, Lowrey & Co.

Payne, Arvilla C. 1980. Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In William Labov (ed.), *Locating language in time and space*, 179–218. New York: Academic Press.

Peirce, Jonathan W. 2007. PsychoPy – psychophysics software in Python. *Journal of Neuroscience Methods* 162(1–2). 8–13. http://www.psychopy.org/PsychoPyManual.pdf.

Pepperberg, Irene M. 1981. Functional vocalizations by an African Grey parrot (sittacus erithacus). *Zeitschrift für Tierpsychologie* 55(2). 139–160. https://doi.org/10.1111/j.1439-0310.1981.tb01265.x.

Pepperberg, Irene M. 2007. Grey parrots do not always 'parrot': the roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Sciences* 29(1). 1–13. https://doi.org/10.1016/j.langsci.2005.12.002.

Phillips, Betty S. 2001. Lexical diffusion, lexical frequency, and lexical analysis. In J. Bybee & P.J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 123–126. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.07phi.

Phillips, Jacob B. 2018. Phonological environment and the social perception of American English sibilants. In *Proceedings of the 44th meeting of the berkeley linguistics society*, 243–256. University of California–Berkeley.

Phillips, Jacob B. under review. Temporal dynamics of /s/-retraction in American English .

Pickering, Martin J. & Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(02). https://doi.org/10.1017/S0140525X04000056.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P.J. Hopper (eds.), *Frequency effects and the emergence of lexical structure*, 137–157. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.08pie.

Pinget, Anne-France. 2015. *The actuation of sound change*. Utrecht: LOT. http://www.lotpublications.nl/the-actuation-of-sound-change.

Pleck, Joseph H., Freya L. Sonenstein & Leighton C. Ku. 1993. Masculinity ideology: Its impact on adolescent males' heterosexual relationships. *Journal of Social Issues* 49. 1–29. https://doi.org/10.1111/j.1540-4560.1993.tb01166.x.

Podesva, Robert J. & Janneke Van Hofwegen. 2014. How conservatism and normative gender constrain variation in Inland California: The case of /s/. In *University of Pennsylvania Working Papers in Linguistics*, vol. 20 2, 129–137. Philadelphia. http://repository.upenn.edu/pwpl/vol20/iss2/15.

Provine, Robert R. 1989. Contagious yawning and infant imitation. *Bulletin of the Psychonomic Society* 27(2). 125–126. https://doi.org/10.3758/BF03329917.

R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna. http://www.R-project.org/.

Read, Charles. 1975. Children's categorization of speech sounds in English. In *NCTE research report*, vol. 17, Urbana, IL: National Council of Teachers of English.

Reinisch, Eva & Matthias J. Sjerps. 2013. The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics* 41(2). 101–116. https://doi.org/10.1016/j.wocn.2013.01.002.

Reiterer, Susanne M., Xiaochen Hu, T. A. Sumathi & Nandini C. Singh. 2013. Are you a good mimic? neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology* 4. 782. https://doi.org/10.3389/fpsyg.2013.00782.

Repp, Bruno H. 1981. Two strategies in fricative discrimination. *Perception & Psychophysics* 30(3). 217–227. https://doi.org/10.3758/BF03214276.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini & Jiahong Yuan. 2011. *FAVE (Forced Alignment and Vowel Extraction) Program Suite*. http://fave.ling.upenn.edu/FAAValign.html.

Rutter, Ben. 2011. Acoustic analysis of a sound change in progress: The consonant cluster /stɪ/ in English. *Journal of the International Phonetic Association* 41. 27–40. https://doi.org/10.1017/S0025100310000307.

Sachs, Jacqueline, Philip Lieberman & Donna Erickson. 1973. Anatomical and cultural determinants of male and female speech. In Roger W. Shuy & Ralph W. Fasold (eds.), *Language attitudes: Current trends and prospects*, 74–85. Washington, D.C.: Georgetown University Press.

Scarborough, Rebecca. 2004. *Coarticulation and the structure of the lexicon*. Los Angeles: University of California, Los Angeles dissertation. http://phonetics.linguistics.ucla.edu/research/Scarb_diss.pdf.

Scarborough, Rebecca. 2013. Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics* 41(6). 491–508. https://doi.org/10.1016/j.wocn.2013.09.004.

Schober, Michael F & Herbert H Clark. 1989. Understanding by addressees and overhearers. *Cognitive Psychology* 21(2). 211 – 232. https://doi.org/10.1016/0010-0285(89)90008-X.

Shadle, Christine & Sheila Mair. 1996. Quantifying spectral characteristics of fricatives. In *Proceedings of the 4th International Conference on Spoken Language Processing*, https://doi.org/10.1109/ICSLP.1996.607906.

Shadle, Christine H. 2012. On the acoustics and aerodynamics of fricatives. In Abigail C. Cohn, Cécile Fougeron & M. K. Huffman (eds.), *The oxford handbook of laboratory phonology*, 511–526. https://doi.org/10.1093/oxfordhb/9780199575039.001.0001.

Shapiro, Michael. 1995. A case of distant assimilation: /str/ → /ʃtr/. *American Speech* 70(1). 101–107. https://doi.org/10.1215/00031283-75-1-82.

Shepard, C, Howard Giles & B Poired. 2001. Communication accommodation theory. In W. Peter Robinson & Howard Giles (eds.), *The new handbook of language and social psychology*, 33–56. New York: Wiley.

Shockley, Kevin, Laura Sabadini & Carol A. Fowler. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66. 422–429. https://doi.org/10.3758/BF03194890.

Sievers, Eduard. 1901. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der Indogermanischen Spachen*. Leipzig: Breitkopf & Härtel.

Smith, Bridget J, Jeff Mielke, Lyra Magloughlin & Eric Wilbanks. 2019. Sound change and coarticulatory variability involving english /ɹ/. *Glossa* 4(1). 63. https://doi.org/10.5334/gjgl.650.

Solé, Maria-Josep. 2007. Controlled and mechanical properties in speech:a review of the literature. In M.-J. Solé, P. Beddor & M. Ohala (eds.), *Experimental approaches to phonology*, 302–321. Oxford: Oxford University Press.

Sonderegger, Morgan. 2012. *Phonetic and phonological dynamics on reality television*. Chicago: University of Chicago dissertation.

Sonderegger, Morgan, Max Bane & Peter Graff. 2017. The medium-term dynamics of accents on reality television. *Language* 93(3). 598–640. https://doi.org/10.1353/lan.2017.0054.

Stevens, Kenneth. 1971. Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America* 50. 1180–1192. https://doi.org/10.1121/1.1912751.

Stevens, Kenneth. 1988. *Acoustic phonetics*. Cambridge, Mass.: MIT Press.

Stevens, Kenneth & Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38(1). 10–19. https://doi.org/10.1016/j.wocn.2008.10.004.

Stevens, Mary & Jonathan Harrington. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58. 118–134. https://doi.org/10.1016/j.wocn.2016.08.003.

Stevens, Mary, Jonathan Harrington & Florian Schiel. 2019. Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa* 4(1). 8. https://doi.org/10.5334/gjgl.620.

Stewart, Mary E. & Mitsuhiko Ota. 2008. Lexical effects on speech perception in individuals with "autistic" traits. *Cognition* 109(1). 157–162. https://doi.org/10.1016/j.cognition.2008.07.010.

Strand, Elizabeth A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1). 86–100. https://doi.org/10.1177/0261927X99018001006.

Stuart-Smith, Jane. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In Jennifer Cole & José Ignacio Hualde (eds.), *Laboratory phonology*, vol. 9, 65–86. New York: Mouton de Gruyter.

Studdert-Kennedy, Michael. 2000. Imitation and the emergence of segments. *Phonetica* 57(2-4). 275–283. https://doi.org/10.1159/000028480.

Studdert-Kennedy, Michael. 2005. How did language go discrete? In M Tallerman (ed.), *Language origins: Perspectives on evolution*, 48–67. Oxford University Press.

Toscano, Joseph C. & Bob McMurray. 2012. Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception & Psychophysics* 74(6). 1284–1301. https://doi.org/10.3758/s13414-012-0306-z.

Treiman, Rebecca. 1985. Phonemic awareness and spelling: Children's judgments do not always agree with adults'. *Journal of Experimental Child Psychology* 39(1). 182–201. https://doi.org/10.1016/0022-0965(85)90035-9.

Trudgill, Peter. 1981. Linguistic accommodation: Sociolinguistic observations on a sociopsychological theory. In R Hendrick, C Masek & M F Miller (eds.), *Papers from the parasession on language and behavior*, Chicago: Chicago Linguistics Society.

Trudgill, Peter. 1986. *Dialects in contact*. New York: Blackwell.

Trudgill, Peter. 2008. Colonial dialect contact in the history of European languages: On the irrelevance of identity in new-dialect formation. *Language in Society* 37. 241–280. https://doi.org/10.1017/S0047404508080287.

Warren, Paul. 2006. /s/-retraction, /t/-deletion and regional variation in New Zealand English /str/ and /stj/ clusters. In P. Warren & C. Watson (eds.), *11th Australian International Conference on Speech Science & Technology*, https://assta.org/proceedings/sst/2006/sst2006-138.pdf.

Webb, J. T. 1970. Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In A. W. Siegman & B. Pope (eds.), *Studies in dyadic communication: Proceedings of a research conference on the interview*, 115–133. New York: Pergamon.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In W. Lehmann & A. Malkiel (eds.), *Directions for historical linguistics*, University of Texas Press.

Whalen, D. H. 1990. Coarticulation is largely planned. *Journal of Phonetics* 18. 3–35. https://doi.org/10.1016/S0095-4470(19)30356-0.

Whalen, D. H. 1991. Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics* 50(4). 351–360. https://doi.org/10.3758/BF03212227.

Wilbanks, Eric. 2017. Social and structural constraints on a phonetically motivated change in progress: (str) retraction in Raleigh, NC. In *Penn working papers in linguistics*, vol. 23 1, http://repository.upenn.edu/pwpl/vol23/iss1/33.

Young, S. J. 1994. *The HTK Hidden Markov Model Toolkit: Design and philosophy*. Entropic Cambridge Research Laboratory, Ltd. http://htk.eng.cam.ac.uk/.

Yu, Alan C. L. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE* 5(8). e11950. https://doi.org/10.1371/journal.pone.0011950.

Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*, 201–227. Oxford: Oxford University Press.

Yu, Alan C. L., Carissa Abrego-Collier, Jacob Phillips, Betsy Pillion & Daniel Chen. 2015. Investigating variation in English vowel-to-vowel coarticulation in a

longitudinal phonetic corpus. In *Proceedings of the 18th International Congress of Phonetic Sciences*, https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0519.pdf.

Yu, Alan C L, Carissa Abrego-Collier & Morgan Sonderegger. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE* 8(9). e74746. https://doi.org/10.1371/journal.pone.0074746.

Zellou, Georgia, Delphine Dahan & David Embick. 2017. Imitiation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience* 1–16. https://doi.org/10.1080/23273798.2016.1275710.

Zellou, Georgia, Rebecca Scarborough & Kuniko Nielsen. 2016. Phonetic imitation of coarticulatory vowel nasalization. *Journal of the Acoustical Society of America* 140(5). 3560–3570. https://doi.org/10.1121/1.4966232.

Zellou, Georgia & Meredith Tamminga. 2014. Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics* 47. 18–35. https://doi.org/10.1016/j.wocn.2014.09.002.

Zimman, Lal. 2013. Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men. *Journal of Language and Sexuality* 2. 1–39. https://doi.org/10.1075/jls.2.1.01zim.

Zimman, Lal. 2017. Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/. *Langauge in Society* 46(3). 339–370. https://doi.org/10.1017/S0047404517000070.