

**/tɹ/ and /dɹ/ in  
North American English:**

Phonologization of a Coarticulatory Effect

**Lyra Magloughlin**

A thesis presented to the  
Faculty of Graduate and Postdoctoral Studies  
in partial fulfillment of the requirements for the  
Doctorate in Philosophy degree in Linguistics



**uOttawa**

Department of Linguistics  
Faculty of Arts  
University of Ottawa

# Abstract

This dissertation examines an observed sound pattern in North American English: the affrication of /t/ and /d/ before /ɪ/. This pattern, which has been reported in many dialects of English, results in words like *tree* and *dream* pronounced more like [tʃi:] and [dʒi:m]. To the best of my knowledge, there have been no experimental investigations into this phenomenon. This research aims to fill that gap.

The first phase of research involved conducting an Apparent Time Study, with the goal of answering the question (**Q1**): Is the phenomenon of /tɪ/ and /dɪ/ affrication in English a sound change in progress? During this initial phase of research, affrication of /tɪ/ and /dɪ/ was investigated as a possible sound change in progress using acoustic data from the Raleigh corpus (Dodsworth and Kohn, 2012) – a collection of sociolinguistic interviews conducted with English speakers born and raised in Raleigh, North Carolina.

The second phase of research, a Production Study conducted in the North Carolina State University Phonology Lab in Raleigh, was designed to examine the interaction between articulators during participants' productions of /t/ and /d/ before /ɪ/, and /t d tʃ dʒ/ before vowels. The goal was to answer the following question (**Q2**): Are English /t/ and /d/ in /tɪ/ and /dɪ/ clusters articulated like prevocalic [t] and [d], like prevocalic [tʃ] and [dʒ], like neither, or like both? This is an important question. Are participants producing [t]s and [d]s coarticulated with [ɪ] that, e.g., for aerodynamic reasons, simply *sound* like affricates, or have participants adopted

a new articulatory target before /ɹ/, suggesting phonologization of a coarticulatory effect?

The final phase of research, the Perception Study, conducted in the same lab, was designed to explore how listeners categorize affricated variants of /t/ before /ɹ/. It was reasoned that if speakers are producing affricated variants, and listeners are perceiving that affrication, listeners may categorize the variant as CH rather than T in a forced-choice spelling task. The Perception experiment consisted of a series of non-words because affricates in complex onsets are prohibited in English. As a result, English has no contrastive /tɹ/ - /tʃɹ/ sequences, meaning words produced as [t.ɪ] and [tʃ.ɪ] are both likely to be categorized as *tree* by listeners, while non-words [t.ɹʌv] and [tʃ.ɹʌv] may be categorized differently. The Perception Study attempts to answer the question (**Q3**): Do English speakers categorize affricated variants of /t/ found in /tɹ/ clusters as T or CH?

Results showed that a sound change was underway for English speakers in Raleigh, North Carolina by the middle of the 20th century (Q1), that young English speakers were producing distinct articulatory targets for /t/ and /d/ in /tɹ/ and /dɹ/ sequences, and not simply [t]s and [d]s coarticulated with [ɹ] (Q2), and that these same speakers were categorizing affricated variants of /t/ as CH when taken outside of their natural phonetic environment, with some categorizing affricated variants of /t/ and phonological /tʃ/ similarly in pre-[ɹ] environments (Q3). Taken together, a clearer picture of the phenomenon of /tɹ/ and /dɹ/ affrication has begun to emerge.

# Dedication

To my daughter, Zoë. This is for you, munchkin. I promise never to write another “stupid dissertation” (as you once called this one), or any other kind of dissertation, ever again. I am so proud of you.

To my mother, Cynthia (aka Grandma). This is also for you, Mom. Thank you for your tremendous support, faith, encouragement, patience, and love. I would not have gotten here without you.



Zoë, age 4: “Mummy and ultrasounds”

# Acknowledgements

Thank you to my supervisor Jeff Mielke for always challenging me, and pushing me to keep puzzling through problems, even when I was convinced they'd gotten the best of me. I have learned an astonishing amount from Jeff, but will value most having understood the importance of allowing my brain time to reason things through. I am fortunate to have had the opportunity to work with Jeff, for the depth of his theoretical and experimental knowledge, his insight and ability to see things from perspectives that others do not, his research integrity, and his commitment to his students. I am a better researcher and critical thinker because of Jeff, and for that, I am truly thankful.

I am very grateful to my committee members, Marc Brunelle, Ian MacKay, Kevin McMullin, and Meredith Tamminga. The final version of this dissertation is stronger thanks to their insightful feedback, suggestions, and questions, which prompted very good post-defence discussions (and the inclusion of two new figures in Chapter 2!).

Thank you to Walt Wolfram and the folks at NCSU for allowing me to visit for 10 months and for making me feel welcome (and thanks Jeff, for inviting me). I am grateful to Robin Dodsworth for her kindness, knowledgeableness, and support, and for creating and sharing her Raleigh Corpus, and Erik Thomas for many thought-provoking discussions and help with stimuli creation. This research has benefited directly from my contact with Eric Wilbanks, whose input and technical assistance during the early stages of the project were invaluable, and whose parallel research on s-retraction in Raleigh led to some interesting collaborative work. Thanks also to Bridget Smith for several useful discussions, to Amanda Eads, Antoun Chahoud, and Amy Hemmeter, for being wonderful, and to Jackson Buck, Paige Duffy, Amanda Eads, Michael J. Fox, Jessica Hatcher, Bridget Smith, Eric Wilbanks, Elizabeth Wright, and Laura Zdanski for helping with the data (in various ways). And thanks to the UNC Chapel Hill P-Side Group: Jennifer Smith, Emily Moeng, Amy Reynolds, and others, for many interesting discussions.

Our time in Raleigh was magical, and we were blessed to be surrounded by people who made us feel at home. Living next door to chickens was a definite highlight and I am sure the shrew population is doing better since our cat (Henry) moved out.

But my real academic home for these past many years was the University of Ottawa.

I am grateful to Ana Arregui for her calming kindness and abundant reasonableness. There would be no dissertation without Ana. This is not hyperbole. I am indebted to Maurice Belanger – a technology genius and an extraordinary human being. Thank you to Rob Truswell for his mentorship (I am so happy Z and I got the chance to know him, Sarah, and Sean and I hope our paths will cross again one day soon). Laura Sabourin and Tania Zamuner kept me on the straight and narrow more times than I can count, and gave me very good advice when I needed it most (thank you!), and Éric Mathieu gave me funding to run a clay tongue modelling workshop, and told me that, after the PhD, everything else would be comparatively easy (words I clung to along the way). Thanks also to the administrative team (without whom, nothing would ever get done): Donna Desbiens, Jeanne D’Arc Turpin, Suzanne Dalrymple, Christine Bertrand, Solange Castillo, and Danielle O’Connor.

I could not have asked for a better cohort, Abdul Alamri, Saleh AlQahtani, Michele Burkholder, and Tharanga Weerasooriya, or more fabulous office-mates, William Dalton, Maxime Papillon, Kate Riccomini, Nova Starr, and Ray Therrien. Our office was where Nova, Kate, and I created the **Z**one of **P**roductivity. I am thankful for these friendships, and for a number of others: Gita Zareikar (my PhD sister) and I kept pace with one another (defending mere days apart) and experienced the highs and lows of dissertating together. I can only hope that she felt as much encouragement and support from me as I did from her. Thank you also to Brandon Fry for pep-talks over drinks, Sawsan Alwabari for talking shop with me, and Sonia Pritchard, Jumanah Abdulwahab Abusulaiman, Reza Falahati, Laura Kasronic, Joy Li, Jonathan Lim, Nicté Fuller Medina, Vesela Simeonova, and Félix Desmeules-Trudel, my fellow-travellers.

My love for linguistics was kindled as an undergraduate student at the University of Victoria, and I am thankful to Dave McKercher (for his mentorship), Tae-Jin Yoon, Sonya Bird, John Esling, Janet Leonard, and Isabelle Grenon (who taught me how to write competitive funding applications). I was lucky to spend a year as a visiting undergraduate student at Carleton University, and I have continued to benefit from my interaction with Lev Blumenfeld (who encouraged me to pursue graduate studies in linguistics – thank you for many interesting discussions, hikes, and glasses of wine), Dan Siddiqi, Ida Toivonen, Jaffer Sheyholislami, Beth MacLeod, and Karen Jesney. This research has also benefited from feedback at NWAV45, LabPhon15, a colloquium talk at Carleton University, and discussions with Bryan Gick, and Meghan Clayards.

To my friends outside of academia, I would have been lost without you. Giovanni Cotroneo, I am thankful to you for so many things, but mostly, for keeping me calm, believing in me, and giving me a quiet place to write. Ti amo. Billy Barlow, where do I even begin? Thank you for being my best friend forever. We’ll make it to the ocean one day... Karen Noganosh, you continue to amaze and inspire me and I have been grateful for our friendship for many years, and will be for many years to come. And James Holcroft, thank you for believing in me before I believed in myself.

Thank you to my family for their support and encouragement throughout this process. To my mother for stepping in when I was in dissertation hell – Z is very lucky to have you, and so am I. To my father (who prefers to remain anonymous), thank you for being my Dad. To my sisters, Sam and Amara, you two are the funniest people I have ever met – thank you for keeping me laughing, and for your support at various critical points along the way. I love you all.

Finally, thank you to my amazing daughter Zoë. When I broached the idea with my (then) 8-year-old of packing up and moving to North Carolina so I could collect data, she signed on enthusiastically, and drew me a picture entitled: “The Adventures of Lyra and Zoe!”. There have been so many moments when I have been astounded by her strength, resilience, creativity, courage, and adaptability, and I can’t count the number of encouraging cards and inspirational notes she has written to me during this process. She has earned this degree as much as I have, and I can’t wait for us to begin our next adventure...

This research was supported by the Social Sciences and Humanities Research Council of Canada, the Council of Ontario Universities, and the University of Ottawa Faculty of Graduate and Postdoctoral Studies. Data collection and analysis in North Carolina were supported by funding from NSF Grant No. BCS-1451475 “Phonological implications of covert articulatory variation” to Jeff Mielke, and also the Department of English, College of Humanities and Social Sciences, and Provost’s Professional Experience Program at North Carolina State University. I gratefully acknowledge their financial contributions to this project.



Social Sciences and Humanities  
Research Council of Canada

Conseil de recherches en  
sciences humaines du Canada

Canada



uOttawa

Faculté des études  
supérieures et postdoctorales

Faculty of Graduate  
and Postdoctoral Studies



**NC STATE**  
**UNIVERSITY**

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                       | <b>1</b>  |
| 1.1      | Sound Change . . . . .                                    | 3         |
| 1.2      | Phonologization or Contact? . . . . .                     | 5         |
| 1.3      | Coarticulation or a New Target? . . . . .                 | 6         |
| 1.4      | Structure of the Dissertation . . . . .                   | 9         |
| <b>2</b> | <b>Apparent Time Study</b>                                | <b>12</b> |
| 2.1      | Methodology . . . . .                                     | 12        |
| 2.1.1    | Automatic Classification Using Forced Alignment . . . . . | 14        |
| 2.1.1.1  | A(ffrication)-Score . . . . .                             | 17        |
| 2.1.1.2  | A(ffrication)-Ratio . . . . .                             | 18        |
| 2.2      | A-Score Results . . . . .                                 | 19        |
| 2.3      | A-Ratio Results . . . . .                                 | 22        |
| 2.3.1    | Statistical Modeling . . . . .                            | 24        |
| 2.4      | Phonologization or Contact? (Revisited) . . . . .         | 25        |
| 2.5      | Discussion . . . . .                                      | 30        |
| <b>3</b> | <b>Production Study</b>                                   | <b>32</b> |
| 3.1      | Methodology . . . . .                                     | 36        |
| 3.1.1    | Participants . . . . .                                    | 36        |
| 3.1.2    | Stimuli . . . . .   | 37        |
| 3.1.3    | Procedures . . . . .                                      | 38        |
| 3.1.3.1  | Initial set-up . . . . .                                  | 39        |
| 3.1.3.2  | Recording . . . . .                                       | 44        |
| 3.1.4    | Data Processing and Analysis . . . . .                    | 45        |
| 3.1.4.1  | Acoustic analysis . . . . .                               | 45        |
| 3.1.4.2  | Articulatory analysis . . . . .                           | 46        |
| 3.2      | Production Results . . . . .                              | 48        |
| 3.2.1    | Lip rounding . . . . .                                    | 49        |
| 3.2.2    | Tongue contours . . . . .                                 | 57        |
| 3.2.2.1  | Tongue contour analysis . . . . .                         | 58        |
| 3.2.2.2  | Front vowel contexts . . . . .                            | 59        |
| 3.2.2.3  | Back vowel contexts . . . . .                             | 64        |
| 3.2.2.4  | Summary of observed patterns . . . . .                    | 69        |
| 3.2.2.5  | Comparing across tongue curves . . . . .                  | 72        |

|          |  |            |
|----------|--|------------|
| 3.2.3    | Duration . . . . .   | 80         |
| 3.2.3.1  | Statistical modeling . . . . .   | 82         |
| 3.2.4    | Comparing Tongue Contour and Duration Results . . . . .                | 87         |
| 3.3      | Discussion . . . . .   | 89         |
| 3.3.1    | Answering Question 2 . . . . .   | 93         |
| <b>4</b> | <b>Perception Study</b>  | <b>95</b>  |
| 4.1      | Methodology . . . . .  | 95         |
| 4.1.1    | Experimental Design . . . . .  | 96         |
| 4.1.2    | Stimuli . . . . .  | 97         |
| 4.1.2.1  | Construction of Stimuli . . . . .                                      | 100        |
| 4.1.2.2  | Experimental Trials . . . . .  | 103        |
| 4.1.3    | Participants . . . . .   | 104        |
| 4.1.4    | Procedures . . . . .   | 104        |
| 4.2      | Perception Results . . . . .   | 110        |
| 4.2.1    | Overview of Results . . . . .  | 112        |
| 4.2.2    | Comparing Across Categories . . . . .                                  | 116        |
| 4.3      | Discussion . . . . .   | 121        |
| <b>5</b> | <b>Comparing Results Across Studies</b>                                | <b>123</b> |
| 5.1      | Comparing Production and Apparent Time . . . . .                       | 123        |
| 5.2      | Comparing Perception and Production . . . . .                          | 124        |
| 5.2.1    | CH <sub>V</sub> V-T <sub>R</sub> V by Duration Ratio . . . . .         | 125        |
| 5.2.2    | CH <sub>V</sub> V-T <sub>R</sub> V by Tongue Tip/Blade . . . . .       | 127        |
| 5.2.3    | CH <sub>V</sub> R-T <sub>R</sub> R by Duration Ratio . . . . .         | 128        |
| 5.2.4    | CH <sub>V</sub> R-T <sub>R</sub> R by Tongue Tip/Blade Ratio . . . . . | 129        |
| 5.2.5    | Discussion . . . . .   | 130        |
| <b>6</b> | <b>General Discussion</b>  | <b>134</b> |
| 6.1      | Conclusion . . . . .   | 143        |
| <b>A</b> | <b>Experimental Stimuli</b>  | <b>145</b> |
| A.1      | Production Stimuli and Filler . . . . .                                | 145        |
| A.2      | Perception Non-Word Stimuli and Filler . . . . .                       | 147        |
| <b>B</b> | <b>Consent Form and Questionnaire</b>                                  | <b>149</b> |
| B.1      | University of Ottawa Consent Form . . . . .                            | 149        |
| B.2      | Language Background Questionnaire . . . . .                            | 152        |
|          | <b>Bibliography</b>  | <b>153</b> |

# List of Tables

|     |   |     |
|-----|---|-----|
| 2.1 | Age and sex breakdown of Raleigh corpus speakers . . . . .            | 13  |
| 2.2 | Total number of Raleigh corpus words extracted and included . . . . . | 16  |
| 2.3 | Summary of Model 1a: TR . . . . .                                     | 25  |
| 2.4 | Summary of Model 1b: DR . . . . .                                     | 25  |
| 3.1 | Production Study participant details . . . . .                        | 37  |
| 3.2 | Production stimuli . . . . .  | 38  |
| 3.3 | Summary of observed tongue root patterns . . . . .                    | 70  |
| 3.4 | Summary of observed tongue tip/blade patterns . . . . .               | 71  |
| 3.5 | Summary of Duration Model 1a (TR): by Category . . . . .              | 83  |
| 3.6 | Summary of Duration Model 1b (DR): by Category . . . . .              | 83  |
| 3.7 | Summary of Duration Model 2a (TR): by Category + Vowel . . . . .      | 84  |
| 3.8 | Summary of Duration Model 2b (DR): by Category + Vowel . . . . .      | 84  |
| 4.1 | Perception stimuli . . . . .  | 101 |
| 4.2 | Filler stimuli spelling pairs . . . . .                               | 104 |
| 4.3 | Perception Study participant details . . . . .                        | 105 |
| 4.4 | All participants: Mean CH response rate . . . . .                     | 113 |
| 5.1 | Production-perception comparisons . . . . .                           | 132 |
| A.1 | Production Stimuli . . . . .  | 145 |
| A.2 | Production Filler . . . . .   | 146 |
| A.3 | Perception Stimuli . . . . .  | 147 |
| A.4 | Perception Filler . . . . .   | 148 |

# List of Figures

|      |   |    |
|------|---|----|
| 2.1  | Modified CMU pronouncing dictionaries . . . . .                           | 16 |
| 2.2  | Histograms of A-Scores for all tokens . . . . .                           | 20 |
| 2.3  | A-Scores across apparent time: voiceless phones . . . . .                 | 21 |
| 2.4  | A-Scores across apparent time: voiced phones . . . . .                    | 22 |
| 2.5  | A-Ratios across apparent time: TR and DR . . . . .                        | 23 |
| 2.6  | Kurtosis: speaker A-Scores by phone . . . . .                             | 27 |
| 2.7  | Standard Deviation: speaker A-Scores by phone . . . . .                   | 29 |
|      |   |    |
| 3.1  | Hypothesized gestural score diagrams . . . . .                            | 33 |
| 3.2  | Participant set-up for Production data collection . . . . .               | 40 |
| 3.3  | Visible anatomical features for transducer placement . . . . .            | 42 |
| 3.4  | On-screen messaging about head position relative to baseline . . . . .    | 43 |
| 3.5  | Ultrasound and video images during recording . . . . .                    | 44 |
| 3.6  | Sample raw tongue traces and SSANOVA . . . . .                            | 47 |
| 3.7  | Example of a single time point across modalities . . . . .                | 48 |
| 3.8  | Hypothesized gestural score diagrams (a and b) . . . . .                  | 49 |
| 3.9  | Orbicular Oris (OO) muscle diagram . . . . .                              | 50 |
| 3.10 | S03: labial constriction comparisons, voiceless tokens . . . . .          | 52 |
| 3.11 | S03: labial constriction comparisons, voiced tokens . . . . .             | 53 |
| 3.12 | S23: labial constriction comparisons, voiceless tokens . . . . .          | 54 |
| 3.13 | S09: labial constriction comparisons, voiceless tokens . . . . .          | 55 |
| 3.14 | All participants: start of /ɪ/ in <i>rub</i> . . . . .                    | 56 |
| 3.15 | All participants: start of /ɪ/ in <i>truck</i> . . . . .                  | 56 |
| 3.16 | Hypothesized gestural score diagrams (a and b) . . . . .                  | 58 |
| 3.17 | S05: tongue contours, /ɪ/ vowel context . . . . .                         | 60 |
| 3.18 | S18: tongue contours, /ɪ/ vowel context . . . . .                         | 61 |
| 3.19 | S03: tongue contours, /i/ vowel context . . . . .                         | 62 |
| 3.20 | S30: tongue contours, /i/ vowel context . . . . .                         | 63 |
| 3.21 | S18: tongue contours, /ɑ/ vowel context . . . . .                         | 65 |
| 3.22 | S09: tongue contours, /ʌ/ vowel context . . . . .                         | 66 |
| 3.23 | S24: tongue contours, /ɑ/ vowel context . . . . .                         | 67 |
| 3.24 | S23: tongue contours, /ɑ/ and /ʌ/ vowel contexts . . . . .                | 68 |
| 3.25 | Tongue root comparisons across vowel contexts: mid-closure . . . . .      | 74 |
| 3.26 | Tongue root comparisons across vowel contexts: mid-noise . . . . .        | 75 |
| 3.27 | Tongue tip/blade comparisons across vowel contexts: mid-closure . . . . . | 77 |
| 3.28 | Tongue tip/blade comparisons across vowel contexts: mid-noise . . . . .   | 78 |

|      |   |     |
|------|---|-----|
| 3.29 | S18: concavity at the tongue dorsum across vowel contexts . . . . .             | 79  |
| 3.30 | All participants: noise interval duration . . . . .                             | 81  |
| 3.31 | All participants: noise interval duration, by vowel . . . . .                   | 82  |
| 3.32 | S03: typical noise interval duration . . . . .                                  | 85  |
| 3.33 | S24: atypical noise interval duration . . . . .                                 | 86  |
| 3.34 | All participants: Tongue Tip/Blade Ratio by Duration Ratio . . . . .            | 88  |
| 3.35 | Revised gestural score diagram, based on observed data . . . . .                | 90  |
|      |   |     |
| 4.1  | List of Perception experiment non-words for recording . . . . .                 | 98  |
| 4.2  | Perception experiment stimuli: noise interval duration . . . . .                | 99  |
| 4.3  | Construction of Perception experiment stimuli . . . . .                         | 100 |
| 4.4  | Example of a stop spliced from monosyllabic token . . . . .                     | 102 |
| 4.5  | Example of a CV+stop sequence spliced from multisyllabic token . . . . .        | 103 |
| 4.6  | Perception experiment on-screen instructions . . . . .                          | 107 |
| 4.7  | Practice phase screens . . . . .  | 108 |
| 4.8  | Spelling pairs for experimental stimuli . . . . .                               | 109 |
| 4.9  | All participants: CH categorizations, word-initial position . . . . .           | 112 |
| 4.10 | All participants: CH categorizations, word-medial position . . . . .            | 113 |
| 4.11 | All participants: CH categorizations, word-initial position, by vowel . . . . . | 114 |
| 4.12 | All participants: CH categorizations, word-medial position, by vowel . . . . .  | 115 |
| 4.13 | Comparing CH <sub>v</sub> V and TvV categories . . . . .                        | 116 |
| 4.14 | Comparing TvR and TvV categories . . . . .                                      | 117 |
| 4.15 | Comparing CH <sub>v</sub> R and CH <sub>v</sub> V categories . . . . .          | 118 |
| 4.16 | Comparing TrR and TrV categories . . . . .                                      | 119 |
| 4.17 | Comparing CH <sub>v</sub> V and TrV categories . . . . .                        | 120 |
| 4.18 | Comparing CH <sub>v</sub> R and TrR categories . . . . .                        | 120 |
|      |   |     |
| 5.1  | CH <sub>v</sub> V-Tr <sub>R</sub> V by Duration Ratio . . . . .                 | 127 |
| 5.2  | CH <sub>v</sub> V-Tr <sub>R</sub> V by Tongue Tip/Blade Ratio . . . . .         | 128 |
| 5.3  | CH <sub>v</sub> R-Tr <sub>R</sub> R by Duration Ratio . . . . .                 | 129 |
| 5.4  | CH <sub>v</sub> R-Tr <sub>R</sub> R by Tongue Tip/Blade Ratio . . . . .         | 130 |

# Chapter 1

## Introduction

This dissertation examines an observed sound pattern in North American English: the affrication of /t/ and /d/ before /ɹ/. This pattern, which has been reported in many dialects of English, results in words like *tree* and *dream* pronounced more like [tʃi:] and [dʒi:m]. To the best of my knowledge, there have been no experimental investigations into this phenomenon, yet it is often referred to quite matter-of-factly in the literature. Read (1971:13), for example, describes how, "before [r] in English, [t] and [d] are affricated, i.e., released slowly with a resulting ‘shh’ sound". Smith (2013:71) refers to /tr/ affrication as “a sound change nearly completed”, and while Stevens and Harrington (2016:119) note that “there is surprisingly little discussion of the sound change /tɹ/ > [tʃɹ] in the literature”, they do not question its status as such. In *Gimson’s Pronunciation of English* (Cruttenden, 2014:189–192), foreign learners are warned about possible confusion between /tʃ/ and /tɹ/, and /dʒ/ and /dɹ/ in minimal pairs such as “*cheese, trees*” and “*jaw, draw*”.

The affrication of /tɹ/ has also been implicated in another English phenomenon. Lawrence (2000) asserts that it is the affrication of /t/ that triggers s-retraction, a well-documented sound pattern observed in English dialects across the United States (Baker et al., 2011; Durian, 2007; Gylfadottir, 2015; Labov, 2001; Mielke

et al., 2010; Rutter, 2011; Wilbanks, 2017), the United Kingdom (Bass, 2009; Glain, 2014), and New Zealand (Lawrence, 2000). English /tɹ/ and /dɹ/ affrication have also been reported outside of North America. Gordon and Maclagan (2004:612), for example, refer to the affrication of /tr/ as a rapid change that is occurring in New Zealand English, noting that while “the /t/ in /tr/ has always partially devoiced the following /r/,” now, “the lips are being rounded, and the cluster is pronounced as though it were spelt *chr*, so that *tree* is now pronounced [tʃ̥i:]”. Less recently, Jones (1963:80) observed that “Southern [British] English **tr** and **dr** seem to be intermediate between single affricates and sequences of two distinct sounds”.

Ohala and Solé (2010:47) present a phonetic explanation for what they refer to as “emergent affricates” that result from the extended period of turbulence that occurs as a stop is released into a segment with a high tongue position – a vowel, a glide, or (as in this case) a liquid. In their framing, increased friction noise occurs during the early portion of a segment with narrow constriction, but can be interpreted by a listener as belonging to the preceding stop (stop + prolonged friction noise = affricate), rather than to the following, narrowly constricted segment. This could lead to the phonologization of a coarticulatory effect, if a listener interprets the emergent affricate as a new articulatory target.

Given the assumptions that have been made about English /tɹ/ and /dɹ/ affrication in the literature, it seems somewhat remarkable that there is a complete absence of experimental work investigating the phenomenon. Consequently, there are still many unanswered questions. Is the affrication of /t/ and /d/ before /ɹ/ a sound change in progress? Are the ‘affricates’ simply emergent, as suggested by Ohala and Solé (2010), due to aerodynamic factors, or have some speakers phonologized a coarticulatory effect? This dissertation examines acoustic, articulatory, and perceptual data on /tɹ/ and /dɹ/, in order to address these questions.

## 1.1 Sound Change

The Actuation Problem, a term originally coined by Weinreich, Labov, and Herzog (1968), refers to the unresolved question of why a sound change occurs where and when it does, yet fails to occur elsewhere and at other times. Various explanations for sound change have been proposed over the last century and a half, many of which have been situated in the speaker's drive to reduce articulatory effort (Paul, 1880; Whitney, 1867). In his theory of hyper- and hypo-articulated speech (H&H theory), Lindblom (1990) expanded this to include the listener's needs, arguing that speakers balance their desire to reduce articulatory effort (through hypo-articulated speech) with their attentiveness to the situation-specific needs of the listener (through hyper-articulated speech). Sound change, in this sense, can occur when a speaker miscalculates the listener's needs, hyperarticulating to the point where the listener assumes a different target from the one intended by the speaker, thereby expanding the range of communicative (phonetic) variants available to the listener-turned-speaker (Lindblom et al., 1995). Ohala (1981, 1983, 1989, 1993, 1994) proposes a different explanation for change, asserting that phonetic variability in the speech signal is something to be factored out ('normalized') by the listener, and sound change is something that results from a listener's failure to do so – the phonologization of coarticulation. In *Evolutionary Phonology*, Blevins (2004) presents a typology of sound change. In her typology, CHOICE refers to a listener accurately perceiving a speaker's hyper- and hypo-articulated phonetic variants, but selecting a different phonological target from the speaker's, similar in nature to Lindblom's (1990) H&H theory. CHANGE, in contrast, focuses on listener misperceptions of the phonetic signal, in line with Ohala's work, whereas CHANCE captures instances of inherent ambiguity in the phonetic signal, resulting in different underlying forms for speaker and listener.

What many approaches to sound change have in common is the notion that

change is due to an “accumulation of errors” (Baker, Archangeli, and Mielke, 2011), that result from speaker- and/or listener-based needs. Baker et al. (2011) argue that accumulation-of-error theories such as these are problematic because they over-predict sound change and do not provide an adequate account for when sound change fails to occur. This is the crux of the actuation problem, as framed by Weinreich et al. (1968). In keeping with Janda and Joseph (2003), who argue that the initiation of a sound change is a brief and phonetically sudden event (a ‘burst’), Baker et al. (2011) propose a partial solution to this problem, using s-retraction as the basis for their argumentation. English s-retraction is a phenomenon in which speakers produce a retracted version of /s/, resulting in more [ʃ]-like articulations (e.g. [tʃit] rather than [stɪt]). While s-retraction can occur in a variety of (sC, sCɪ) contexts, it is observed most frequently in /stɪ/ clusters (Shapiro, 1995), where it is at its most extreme (Baker et al., 2011). Baker et al. (2011) observed that /ɪ/ has a depressing effect on the centroid frequency of /s/ (making it more [ʃ]-like) in sCɪ clusters, providing a clear phonetic motivation for retraction. The authors found that while only 8 of the 19 participants in their study were classified as retractors, all participants exhibited “a substantial coarticulatory bias toward s-retraction” (p. 360), particularly in sCɪ contexts. Further, they reported that non-retractors whose /s/ and /ɪ/ tongue shapes exhibited less articulatory distance (i.e. more closely resembled one another), produced /s/ with a lower (more [ʃ]-like) centroid frequency. Variable articulatory behaviour in /ɪ/ production has been documented in both adults (Delattre and Freeman, 1968; Westbury et al., 1998; Mielke et al., 2010) and children (Magloughlin, 2016), with no perceivable acoustic consequences – what Twist et al. (2007) have referred to as ‘covert’ variability. Baker et al. (2011) assert that this covert variability during production can lead to the noticeable exaggeration of a phonetic effect. This is a critical component of the authors’ proposed solution to the actuation problem:

[T]he potential for sound change arises when inter-speaker variability is so great that some speakers actually produce a sound that can be perceived as a distinct target. Given the appropriate social conditions, this potential can be realized through another speaker adopting the novel target in his/her speech. (p. 351)

Baker et al. (2011) argue that the actuation of a sound change is rare because actuation *depends* on “the chance alignment of extreme coarticulation with extreme influence” (*idem.*).

What would the extreme coarticulation of /tɹ/ and /dɹ/ sequences look like? Given the assumptions that have been made in the literature, we might assume (for the time being) that the result could be coarticulated variants that are misperceived as the distinct targets: [tʃɹ] and [dʒɹ]. As affricates in complex onsets are phonotactically restricted in English, [t] and [tʃ] are not contrastive before /ɹ/, so there is no semantic consequence of producing an affricated variant in this context: [ti] and [tʃi] may both be parsed as *tree* (whereas [ti] and [tʃi] mean different things). Gylfadottir (2015) notes that a listener’s awareness of the phonotactics of their language may lead to a tolerance for variation in certain contexts. Listener-tolerated inter-speaker variation during production of /tɹ/ and /dɹ/ may provide the ideal conditions for change. If some speakers produce more affricated variants that are tolerated by some listeners and, as Baker et al. (2011) have proposed, perceived by others as a distinct target, the potential for change exists.

## 1.2 Phonologization or Contact?

Change can happen in a language in a number of different ways. For example, change can occur through dialect contact, when a new variant (that may be an exaggerated version of an already present phonetic effect) is imported into a speech community

from elsewhere, with some speakers adopting the new variant – a change from above (the level of consciousness) (e.g., Labov, 1990). Phonologization, on the other hand, occurs when a phonetic effect is exaggerated within a speech community to the point where it becomes noticeable, at least to some members of the community, and is adopted (phonologized) as a new articulatory target – a change from below.

First introduced by Jakobson (1931), the term phonologization (originally *Phonologisierung*) has been used to refer to a variety of different phenomena over the last century. As noted by Hyman (2013:4), Jakobson’s (1931) meaning “is better translated as phonemicization (whereby an already phonological property changes from allophonic to phonemic)”. In this dissertation, phonologization is used in the sense of Hyman (1977) (revisited in Hyman, 2013), to refer to a phonetic process that has become phonological over time. Hyman (2013:6) proposed two diagnostics for identifying whether phonologization has occurred in a language:

- (i) A phonetic effect is exaggerated beyond what can be considered universal;
- (ii) A ‘categorical’ rule of phonology must refer to the phonologized property.

These diagnostics will be used here to explore whether the affrication of /tɹ/ and /dɹ/ before /ɹ/ is a phenomenon that has been phonologized in the English-speaking community under investigation. Have speakers adopted new articulatory targets that more closely resemble [tʃ] and [dʒ] and, if so, was the change a gradual one, over time, or is the change discontinuous, suggesting that the more affricated variants were imported into the community through dialect contact?

### 1.3 Coarticulation or a New Target?

A key puzzle explored in this dissertation is how to distinguish between a normal amount of coarticulation and a move to a new articulatory target. Teasing these things apart is not necessarily straightforward. What do coarticulated /tɹ/ and

/dɹ/ sequences look like? What is a ‘normal’ (listener-tolerated) amount of coarticulation? In the same way that /k/ may be produced with a more posterior or anterior constriction depending on the adjacent vowel (Frisch and Wodzinski, 2016), are (alveolar) /t/ and /d/ produced with a more posterior constriction when coarticulated with a following (post-alveolar) /ɹ/? Is the aperiodic (frication) noise following the burst of an aspirated coronal stop released into a segment with a high tongue position similar in duration to an affricate? How extreme would a more posterior constriction and/or prolonged aperiodic noise need to be in order for listeners to classify the output as belonging to the category /tʃ/? This is an important question, because it is the existence of the category /tʃ/ in English that provides the opportunity for coarticulation to lead to a change in phonological target – the phonologization of a coarticulatory effect.

The notion that adjacent sounds have an articulatory influence on one another has been around for well over a century (Sweet, 1877; Paul, 1880), and increasingly since the use of experimental methods to study speech sounds (e.g. Rousselot, 1901; Joos, 1948). The term ‘koartikulation’, which was originally coined by Menzerath and de Lacerda (1933), made explicit the idea that adjacent sounds could be articulated together, and not just in sequence with one another: they could be *co-articulated*. Many models of coarticulation have been proposed in the intervening years, with more recent models acknowledging that there are inter-language differences in how sounds are coarticulated - what counts as ‘normal’ coarticulation is language-specific (e.g. Hammarberg, 1976; Guenther, 1994; Beddor, 2009). Feature-based models have drawn criticism (e.g. Fowler, 1980) because they necessitate a ‘translation process’ between abstract, discrete, timeless, invariant phonemes, and the variable and continuous physical realizations of these abstract units during production. This disconnect prompted the development of a number of models that view articulatory gestures as the basic (phonological) units of speech (especially

Browman and Goldstein, 1989), operating under the assumption that sequences of gestures do not alter one another as they transition from one gesture to the next; rather, they are coproduced (e.g. Stetson, 1951; Öhman, 1966; Fowler, 1980; Browman and Goldstein, 1989; Saltzman and Munhall, 1989; Fowler and Saltzman, 1993). Löfqvist (1990:291,316) describes how, “[a]s a consequence of coarticulation, the vocal tract shape at any one time during speech represents an aggregate of gestures associated with different segments”, and argues that the “blending” of these gestures may be advantageous for reasons of “computational simplicity”. This idea has been echoed in work by Guenther (1994), who proposed that speakers actively optimize gestural overlap for reasons of articulatory efficiency, similar to Lindblom’s (1983) ‘economy of effort’. In their task-dynamic model, Fowler and Saltzman (1993:181) suggest that variation in the coproduction of gestures depends on “the degree to which the gestures share articulators”. In cases where there is a conflict between gestures for control of an articulator, the result is intergestural blending. Blending strength refers to the notion that ‘stronger’ gestures are more resistant to coarticulation, but can trigger large coarticulatory effects in ‘weaker’ gestures, which may result in a change in the degree or location of the constriction of the weaker gesture. While Fowler and Saltzman (1993) do not explicitly reference English /ɪ/ in terms of blending strength, their ranking appears to assign greater influence to consonants “requiring extreme constrictions and/or placing strong constraints on articulator movements”. Similarly, the Degree of Articulatory Constraint (DAC) model, introduced by Recasens et al. (1997), posits that lingual coarticulation is determined by the demands placed on the tongue during production. In this model, vowels and consonants are assigned specific DAC values, with those requiring the greatest degree of lingual precision receiving the highest values. Here, a high DAC value means greater coarticulatory resistance and greater coarticulatory influence over adjacent segments.

From the perspective of the models referenced above, it is assumed that /ɹ/ should be classified as a strong gesture in terms of blending strength (Saltzman and Munhall, 1989; Fowler and Saltzman, 1993), and receive a high DAC value (Recasens et al., 1997). While (to the best of my knowledge) no models of coarticulation have looked specifically at English coronal stop + /ɹ/ sequences, we know that /ɹ/ is an articulatorily complex speech sound that requires a high degree of articulatory precision during production. It is also often one of the last sounds to be acquired by children (Sander, 1972; Smit, 1993), particularly in prevocalic position (Stoel-Gammon, 1985; Smit et al., 1990; McGowan et al., 2004; Magloughlin, 2016).

In a recent study comparing the blending strength of the English liquids /ɹ/ and /l/, Walker, Proctor, Smith, and Enzinger (2016) argued that “/ɹ/ in [General American English] is defined by a tongue body gesture with stronger blending parameters<sup>1</sup> than that of the lateral, which results in its exerting a more global influence on tongue shaping and constriction location than /l/”. In keeping with their hypothesis, the authors found that, despite variability in /ɹ/ production, the various tongue postures for /ɹ/ acted to “constrain global tongue shaping more consistently across different positions in the syllable” when compared with /l/ – a finding that may prove relevant when examining the coarticulation of /tɹ/ and /dɹ/. If /ɹ/ exerts more global influence, we may see /t/ and /d/ changing more than /r/.

## 1.4 Structure of the Dissertation

This dissertation argues that the affrication of /t/ and /d/ before /ɹ/ is an active sound change in progress that has been phonologized. Despite numerous references to it in the literature, no experimental work has been undertaken to investigate the phenomenon. This dissertation aims to fill that gap. Conducted over three separate studies, the research presented in subsequent chapters explores three specific

---

<sup>1</sup>For details on blending parameters, see Saltzman and Munhall (1989).

questions:

**Q1: *Apparent Time Study*** – Is the phenomenon of /tɹ/ and /dɹ/ affrication in English a sound change in progress?

**Q2: *Production Study*** – Are English /t/ and /d/ in /tɹ/ and /dɹ/ clusters articulated like prevocalic [t] and [d], like prevocalic [tʃ] and [dʒ], like neither, or like both?

**Q3: *Perception Study*** – Do English speakers categorize affricated variants of /t/ found in /tɹ/ clusters as T or CH?

Chapter 2 presents results from an Apparent Time Study, which examines sociolinguistic interview data from a corpus of Raleigh, North Carolina English speakers of different ages, all born in the 20th century. The Raleigh corpus is considered to be a suitable choice for conducting this investigation for several reasons. First, it is expected that the phonetic motivation for /tɹ/ and /dɹ/ affrication will be present in any English-speaking community. Second, Raleigh experienced an influx of technology workers to the area from Northern regions of the United States in the middle of the 20th century, following the development of Research Triangle Park (RTP), making it plausible that the resulting dialect contact may have led to the introduction of novel affricated variants to the region. Third, /tɹ/ affrication has been implicated in s-retraction (in /stɹ/ clusters), which is a sound change in progress reported in Raleigh English (Wilbanks, 2017). The Apparent Time Study aims to determine whether /tɹ/ and /dɹ/ affrication, if present, is a sound change in progress and/or whether the emergence of affricated variants may have been the result of dialect contact. Building on these findings, Chapter 3 presents results from a Production Study conducted in Raleigh, North Carolina, which captures audio, ultrasound, and video data in order to investigate how English speakers' /tɹ/ and /dɹ/ sequences are coarticulated. The Production Study provides an opportunity

to find out how affricated variants of /t/ and /d/ before /ɹ/ are articulated. Chapter 4 presents results from a Perception Study, which explores how listeners (from the Production Study) categorize affricated variants of /t/ spliced from before /ɹ/. Chapter 5 compares results from across studies, and Chapter 6 provides a general discussion and conclusion.

Ethics approval was obtained through the University of Ottawa Research Ethics Board (REB), and experiments were conducted in compliance with REB laws, regulations, and guidelines.

# Chapter 2

## Apparent Time Study

The first phase of this dissertation research involved conducting an Apparent Time Study, with the goal of answering the question:

**Q1:** Is the phenomenon of /tɹ/ and /dɹ/ affrication in English a sound change in progress?

The apparent time model, a sociolinguistic approach first introduced by Gauchat (1905), is a method of synchronic analysis used to explore change in progress. In contrast with real-time studies, which involve longitudinal data of the same speakers, apparent time studies look at speakers from a range of ages and operate under two assumptions: 1) speech, once it is acquired in childhood, remains relatively stable across the adult lifespan, and 2) speakers from different generations are representative of community-level speech patterns at various points in time (Bailey et al., 1991).

### 2.1 Methodology

During this initial phase of research, affrication of /tɹ/ and /dɹ/ was investigated as a possible sound change in progress using acoustic data from the Raleigh corpus

(Dodsworth and Kohn, 2012) – a collection of sociolinguistic interviews conducted with English speakers born and raised in Raleigh, North Carolina. Speakers included in the analysis consisted of 75 women and 65 men, born between 1923 and 1996 (Table 2.1). Speakers were grouped into three generations, following previous work with the corpus (e.g. Dodsworth and Kohn, 2012). The cut-off between Generation 1 and Generation 2 coincides with the year dialect contact was established in Raleigh following the installation of Research Triangle Park in the mid-1950s, which resulted in an influx of workers from Northern regions of the United States.

| Generation | Birthyear Range | Women | Men | Total      |
|------------|-----------------|-------|-----|------------|
| 1          | 1923-1954       | 28    | 27  | 55         |
| 2          | 1955-1978       | 32    | 24  | 56         |
| 3          | 1979-1996       | 15    | 14  | 29         |
|            |                 | 75    | 65  | <b>140</b> |

**Table 2.1:** Age and sex breakdown of Raleigh corpus speakers

Conversational interviews in the corpus were phonetically aligned using the Penn Phonetics Lab Forced Aligner (P2FA) for English (Yuan and Liberman, 2008). Preliminary acoustic measurements were obtained in Praat (Boersma and Weenink, 2007) using a script created by Wilbanks and Mielke (2015) and adapted by the author. One of the goals at this stage of the research was to develop a robust measure for distinguishing between stops and affricates in the data. This proved challenging. As has been noted by Thomas (2011), “[f]rication noise analysis is a wide-open field” (p. 107). In an apparent time study on s-retraction, Gylfadottir (2015) (following Baker et al., 2011) used centroid frequency to distinguish between /s/ and /ʃ/, and suggested that a similar approach might be used to investigate the affrication of /t/, by measuring the centroid frequency of stop bursts. Smith (2013:63–67) used three

sibilant interval measures to distinguish between different types of affricated and unaffricated variants in /tw/ sequences, including center of gravity (COG), spectral slope, and normalized rise time. Citing Stevens (1998:117–118), Smith (2013:66) noted that “stops have a very short rise time as the constriction is quickly released, while affricates take longer to reach peak intensity”. She also emphasized that stops show more energy at the beginning, in contrast with affricates, which show more energy in the middle. While these types of measures were successful in Smith’s (2013) laboratory data, they proved inadequate for capturing the relevant contrasts in spontaneous speech.

A number of different strategies were employed in an attempt to identify acoustic measures that would distinguish between voiceless prevocalic coronal stops and affricates in the Raleigh corpus data. COG is effective for capturing information about the sibilant interval of fricatives, but for stops and affricates, there are several acoustic events happening in very quick succession. COG velocity was used in order to identify the start and end of the sibilant interval and measures were developed to isolate the maximum COG, normalized rise time (onset of burst to point of maximum intensity divided by duration), intensity of the release burst, sibilant duration, and others. While some of these were able to distinguish between coronal stops and affricates for some speakers, they failed to do so for others, such that no one measure could adequately capture these contrasts across speakers. Instead, an approach was adopted that uses forced alignment to classify speech sounds (as detailed below).

### **2.1.1 Automatic Classification Using Forced Alignment**

Automatic classification using forced alignment is a technique that uses an aligner to identify and assign likelihood scores to segments in a stream of speech. The method makes use of the Penn Phonetics Lab Forced Aligner (P2FA) for English (Yuan and Liberman, 2008), which consists of acoustic models trained with the Hidden

Markov Model Toolkit or HTK (Young et al., 2002), the Carnegie Mellon University (CMU) Pronouncing Dictionary, and a corpus of recordings with over 50 years of oral arguments from the Supreme Court of the United States (SCOTUS). The forced aligner maps word-level transcriptions to audio files and aligns the transcribed words to intervals in the audio file, inserting phone boundaries based on phone-level entries obtained from the pronouncing dictionary. For each force-aligned segment, a likelihood score is calculated based on information from the P2FA acoustic models. Automatic classification involves running the forced aligner more than once using acoustic models from different pronouncing dictionaries, then calculating the difference in likelihood scores for each relevant segment.

Yuan and Liberman (2009) originally developed this automatic classification technique as a way of assigning a darkness score to allophones of /l/, by training new acoustic models categorizing word-initial /l/s as clear, and word-final /l/s as dark. This step was necessary, because P2FA did not include separate acoustic models for the two allophones of /l/. For this dissertation research, it was assumed that affricated versions of /t/ and /d/ would resemble [tʃ] and [dʒ], thereby eliminating the need for a training stage, as the acoustic models for these existing phonemes of English were already present in P2FA.

All non-function words with word-initial /t d tʃ dʒ/ + vowel or /ɹ/ sequences were extracted and force-aligned twice with the P2FA 16 kHz acoustic models, once using a modified version of the CMU Pronouncing Dictionary that contained only pronunciation entries with stops (e.g. TRAIN: [t.ɹem]), and once using a modified version of the dictionary that contained only pronunciation entries with affricates (e.g. TRAIN: [tʃ.ɹem]). An example of these modified dictionaries is presented in Figure 2.1. Words not included in the original CMU Pronouncing Dictionary were added manually to each modified dictionary, as necessary.

Table 2.2 provides a summary of the total number of Raleigh corpus words

extracted and force-aligned across contexts. Twenty speakers were excluded from the analysis for insufficient tokens in one or more contexts (e.g., no /dʒ/ tokens). After exclusions, a total of 14,605 tokens (or 81% of total extracted and force-aligned tokens) were included in the analysis.

| Word Type     | Extracted | Included |
|---------------|-----------|----------|
| #/t/ + /ɹ/    | 3719      | 3082     |
| #/d/ + /ɹ/    | 1883      | 1526     |
| #/t/ + vowel  | 2339      | 1934     |
| #/tʃ/ + vowel | 1017      | 802      |
| #/d/ + vowel  | 7606      | 6128     |
| #/dʒ/ + vowel | 1500      | 1133     |
| TOTAL         | 18064     | 14605    |
|               | 100%      | 81%      |

**Table 2.2:** Total Raleigh corpus words by **Word Type** that were **Extracted** and force-aligned, and the subset of those tokens **Included** in the analysis.

**Figure 2.1:** Modified CMU pronouncing dictionaries. T dict (left) contains exclusively [t] entries for /tɹ/ sequences; CH dict (right) contains exclusively [tʃ] entries for /tɹ/. Similar modified stop/affricate pronouncing dictionaries created for /dɹ/ and /t d tʃ dʒ/ + vowel sequences.

### 2.1.1.1 A(ffrication)-Score

Yuan and Liberman’s (2009) automatic classification technique for classifying /l/ “darkness” resulted in a D(arkness) score for every /l/ token, where a larger D score represented a darker /l/. For this research, Yuan and Liberman’s (2009) technique was adapted in order to calculate an A(ffrication)-Score for every /t d tʃ dʒ/ token before a vowel or an /ɹ/. As shown in Equation 2.1, this was achieved by subtracting the probability (log likelihood score) associated with each segment’s classification as a stop ( $T_1$ ) from the probability of its classification as an affricate ( $T_2$ ), normalized for duration. The higher the A-Score (i.e.,  $> 0$ ), the greater the probability of affrication.

$$(2.1) \quad \mathbf{A}(t) = \log p(t|T_2) - \log p(t|T_1)$$

For example, a Raleigh corpus speaker produced the word *train*. The word was extracted from the corpus because it met the criteria for extraction and was force aligned twice, once as [t.ɹem] (using the T dictionary), and once as [tʃ.ɹem] (using the CH dictionary). P2FA generated a log likelihood score for the /t/ in each alignment: 2209.389161 for the [t.ɹem] alignment ( $T_1$ ), and 2251.318361 for the [tʃ.ɹem] alignment ( $T_2$ ).  $2251.318361 - 2209.389161 = 41.9292$ . An A-Score of greater than 0 suggests that the /t/ in this instance of *train* has a greater probability of being an affricate than a stop, according to the acoustic models of the P2FA. Only A-Scores with values between -100 and +100 were included in the analysis that follows, which represented 96% of tokens.

It is important to note that the P2FA acoustic models for /t/ and /d/ were trained on /t/s and /d/s produced by speakers across a range of contexts, including those before /ɹ/. As a result, the models should be able to accept affricated variants of /t/ and /d/ before /ɹ/ as instances of /t/ and /d/, as long as affricated variants

occurred in the training data. What this means is that affricated variants may be assigned a higher log likelihood score during alignment with the T dictionary (than during alignment with the CH dictionary), which would result in lower (more [t]-like) A-Scores for these tokens than if no affricated variants existed in the training data. However, acoustic models for the P2FA are trained on SCOTUS recordings, which may involve a greater number of older male speakers (Supreme Court justices tend to be old) who may not be producing affricated variants if a change in progress is underway. If fewer affricated variants occurred in the training data, this would increase the likelihood that these affricated variants would derive higher (more [tʃ]-like) A-Scores. Neither scenario is inherently problematic for assessing change over time, given that the same measures are being applied to all tokens in the dataset – an increase or decrease in A-Score can be interpreted as change, regardless of its starting point relative to [tʃ].

#### 2.1.1.2 A(ffrication)-Ratio

In order to measure each speaker’s degree of affrication in /tɹ/ and /dɹ/ productions relative to their prevocalic /t d/ and /tʃ dʒ/ productions, A-Scores were used to calculate A(ffrication)-Ratio. As illustrated in Equation 2.2, A-Ratio was calculated by subtracting a speaker’s mean A-Score for prevocalic /t/ from the A-Score for each /tɹ/ token, and dividing that value by the difference between a speaker’s mean A-Scores for prevocalic /tʃ/ and /t/. This same method was also used to calculate ratios for each /dɹ/ token in the dataset. A ratio closer to 1 indicates that the token is more affricate-like, whereas a ratio closer to 0 indicates that the token is more stop-like. This approach was adapted from Baker et al. (2011), who used median centroid frequency to calculate Retraction Ratio in their investigation of North American /s/-retraction, in order to quantify speakers’ degree of /s/ retraction in sCɹ relative to their prevocalic /s/ (ratio=0) and /ʃ/ (ratio=1). For this data,

a ratio-based approach allows for speaker-internal A-Score normalization: /tɹ/ and /dɹ/ are placed in context within a speaker’s acoustic space, relative to their /t d tʃ dʒ/. In this way, A-Ratio can be used as the dependent measure of /tɹ/ and /dɹ/ affrication.

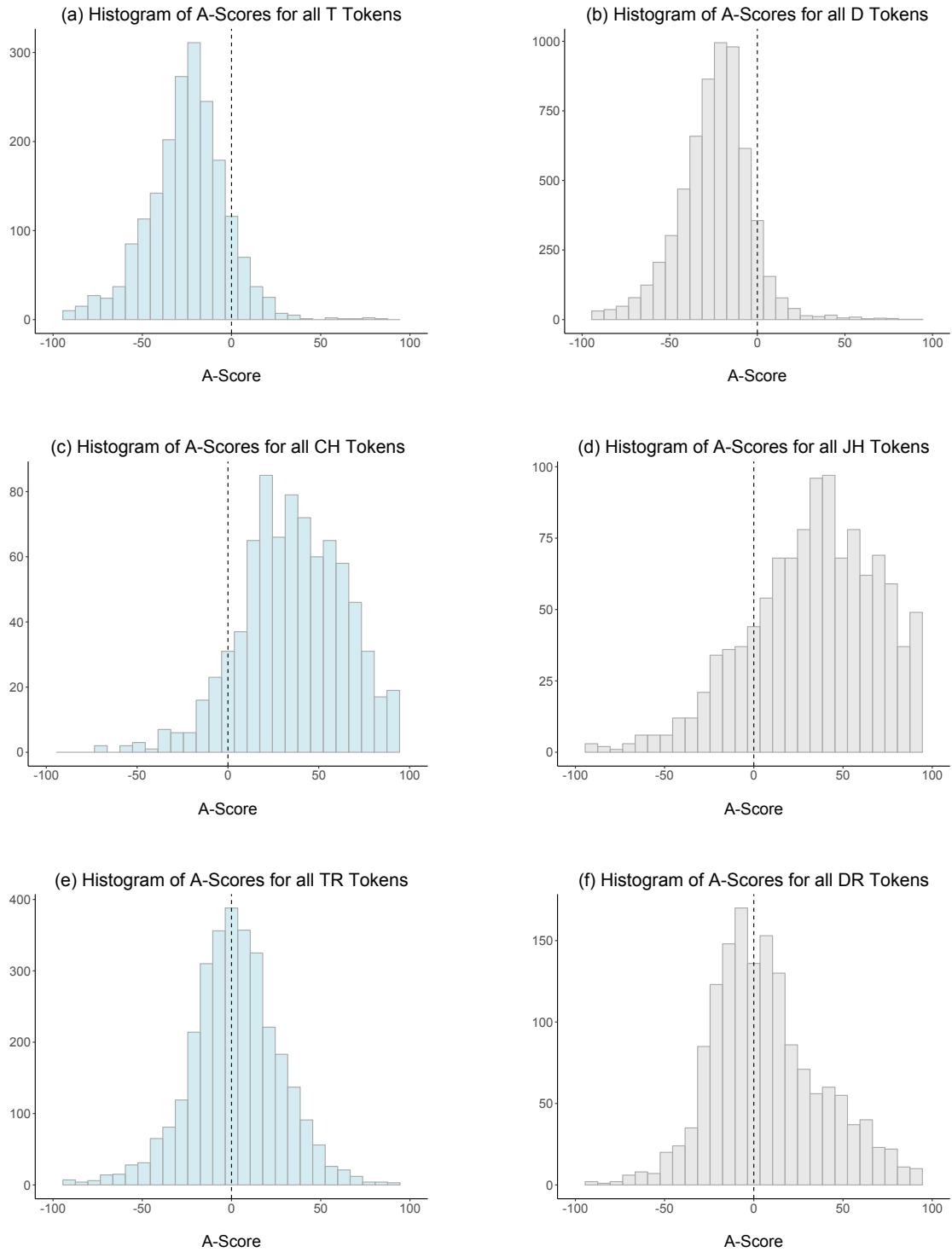
(2.2)

$$\text{A(ffrication)-Ratio} = \frac{\text{(observed A-Score /tɹ/ token — speaker mean A-Score prevocalic /t/)}}{\text{(speaker mean A-Score prevocalic /tʃ/ — speaker mean A-Score prevocalic /t/)}}}$$

## 2.2 A-Score Results

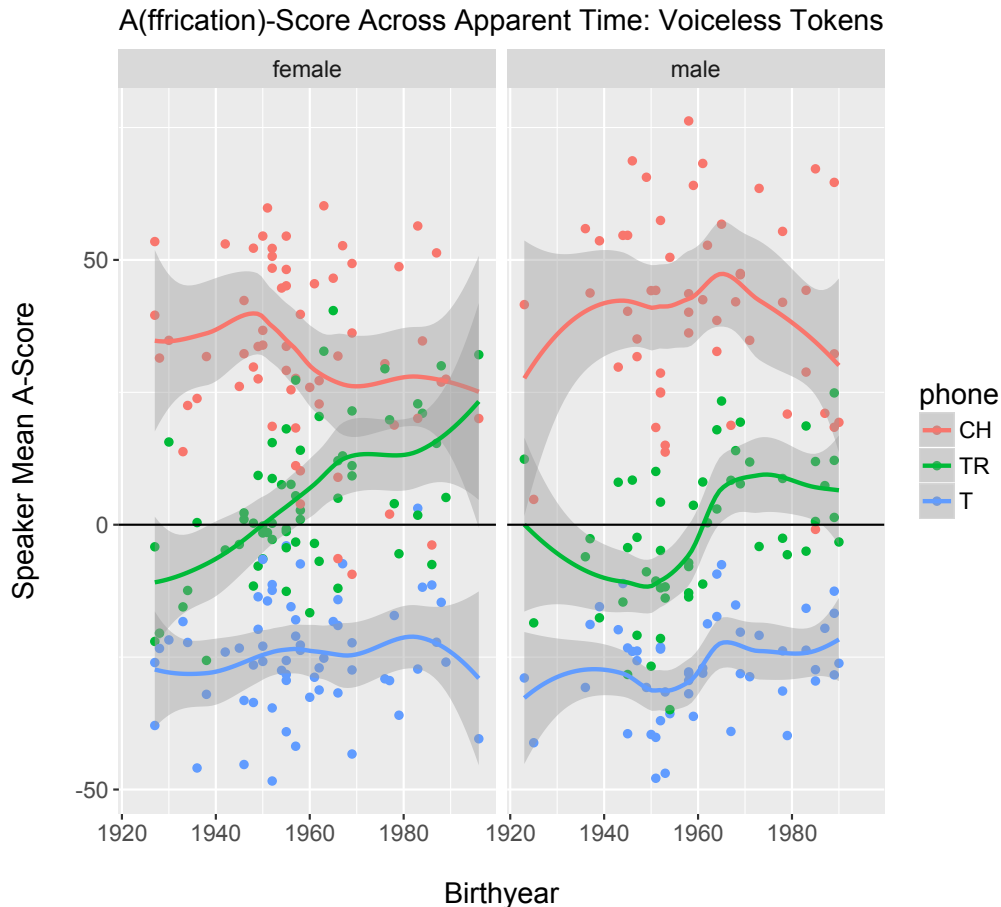
Automatic classification, from which A(ffrication)-Scores were derived, was successful in distinguishing between coronal stops and affricates in the Raleigh corpus data. As described above (§ 2.1.1), A-Scores with values of  $> 0$  indicated that higher log likelihood scores were assigned during forced alignment with the modified affricate pronouncing (CH or JH) dictionary than during forced alignment with the modified stop pronouncing (T or D) dictionary. Figure 2.2 shows predominantly negative A-scores ( $< 0$ ) for prevocalic /t/ and /d/ words (a, b), where no affrication was expected, and predominantly positive A-Scores ( $> 0$ ) for prevocalic /tʃ/ and /dʒ/<sup>1</sup> words (c, d), where affrication was expected. For /tɹ/ and /dɹ/ words (e, f), we see both positive and negative A-Scores, with some tokens showing greater probability of affrication than others, and /dɹ/ showing a greater proportion of tokens with positive A-Scores (relative to negative A-Scores) when compared with /tɹ/.

<sup>1</sup>JH is used by the CMU Pronouncing Dictionary to represent [dʒ], (Figure 2.2, and elsewhere).

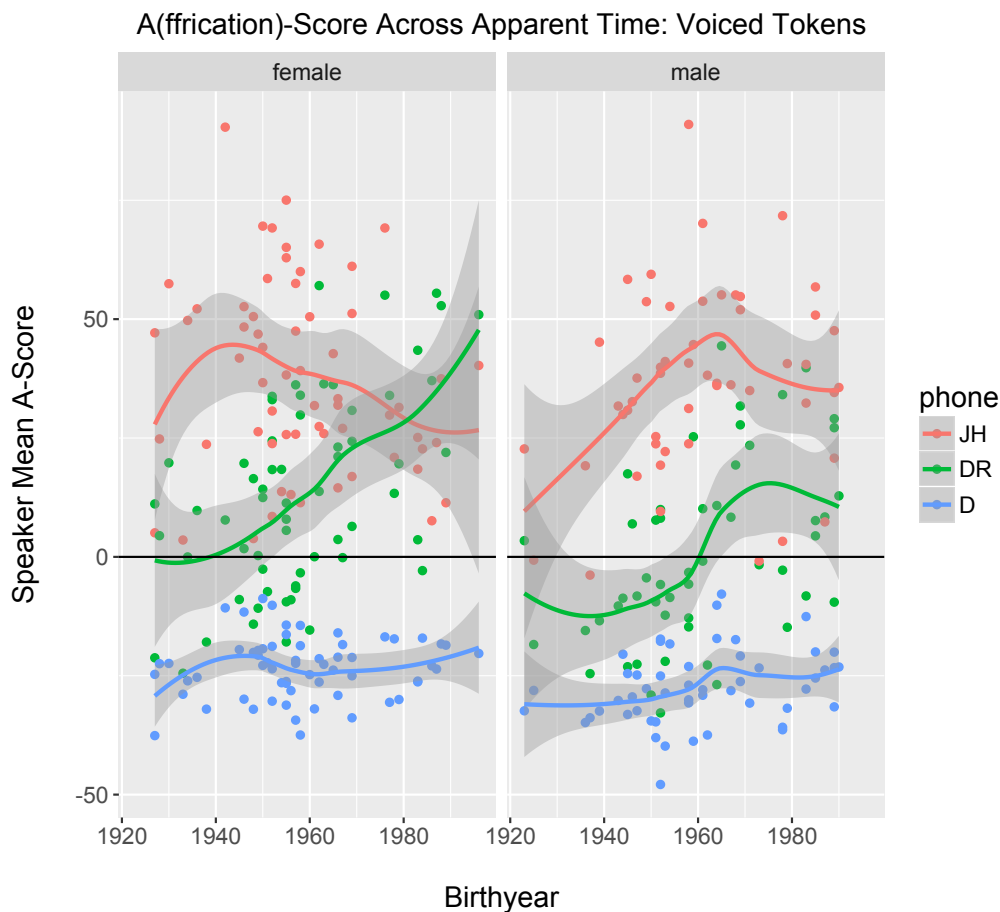


**Figure 2.2:** Histograms of A-Scores for all (a) T and (b) D tokens, (c) CH and (d) JH tokens, and (e) TR and (f) DR tokens. (NOTE: *y*-axis scales for each sub-figure are (at times, dramatically) different, due to variation in token counts - see Table 2.2.)

As illustrated in Figures 2.3 and 2.4, when plotted by birth year, A-Scores for /t/s and /d/s in /tɹ/ and /dɹ/ tokens showed a pattern of change over time for these English speakers in Raleigh, North Carolina. A-Scores were lowest for older male and female speakers, and highest for younger male and female speakers. For voiced tokens (Figure 2.4), the youngest female speakers showed mean A-Score values for DR that were higher than JH, and the oldest male speakers showed mean A-Score values for JH that were much lower relative to the other speakers. These extremes seem largely driven by the comparatively small number of data points (fewer green and orange dots) for females and males in these birth years.



**Figure 2.3:** Mean A-Score of voiceless phones (CH, TR, T) for each female (left) and male (right) speaker by birth year (across apparent time).

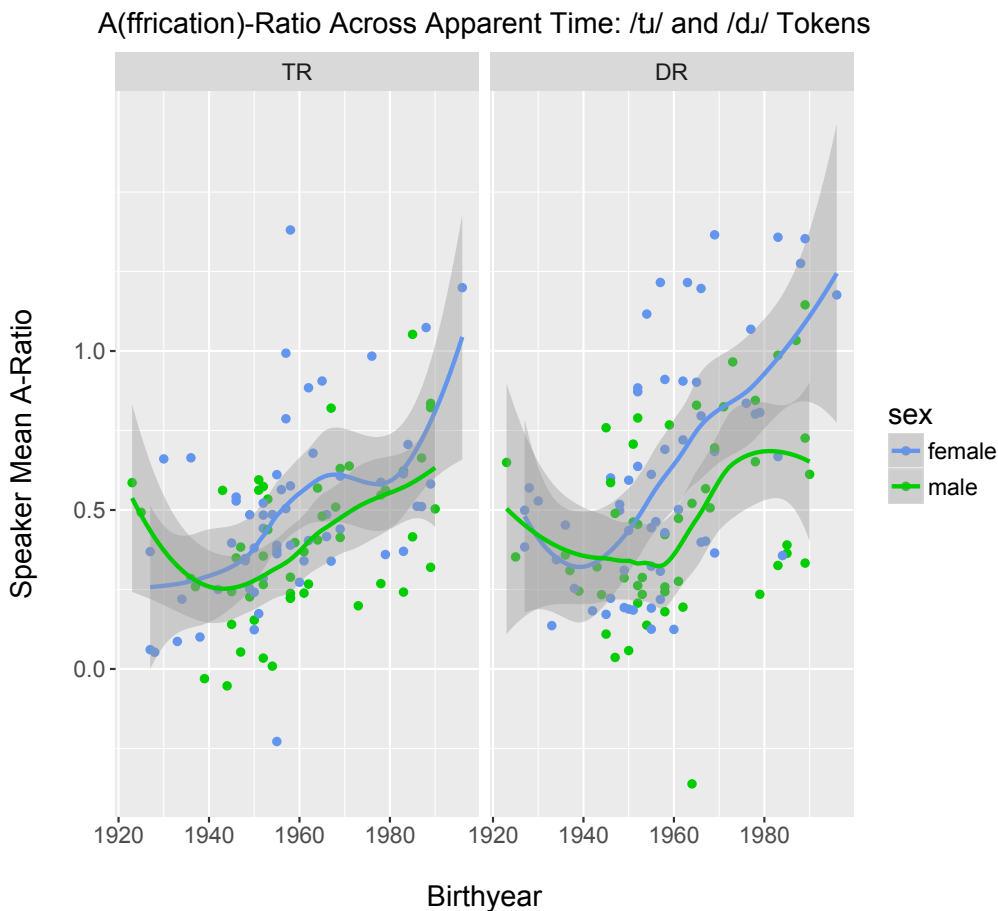


**Figure 2.4:** Mean A-Score of voiced phones (JH, DR, D) for each female (left) and male (right) speaker by birth year (across apparent time).

## 2.3 A-Ratio Results

While raw A-Scores provide information about the likelihood of a speaker’s /tɹ/ or /dɹ/ productions being more or less affricate-like, A(ffrication)-Ratio normalizes these scores by speaker, relative to their prevocalic /t d/ and /tʃ dʒ/ productions. A-Ratio provides information about how far a speaker’s /tɹ/ and /dɹ/ are along the /t/-/tʃ/ and /d/-/dʒ/ dimensions. In keeping with the patterns observed for mean A-Scores, Figure 2.5 shows A-Ratios changing over time. This pattern is most apparent for female speakers, with older females exhibiting mean A-Ratio values closer to 0 for both voiceless (TR) and voiced (DR), and younger females exhibiting

A-Ratio values closer to 1. While a similar pattern can be observed with the males, the change is less apparent, which may be due, in part, to the comparatively small number of data points for the oldest males.



**Figure 2.5:** Mean A-Ratios for /tɹ/ (left) and /dɹ/ (right) for each female and male speaker by birth year (across apparent time).

Strikingly, a number of the youngest male and female speakers show A-Ratios with values even greater than 1, which occurs in cases where a speaker’s mean A-Score for /tʃ/ or /dʒ/ is lower than their mean A-Score for /tɹ/ or /dɹ/. While it is possible that this effect would disappear with more speakers and higher token-counts across categories, it is also plausible that the acoustic information the forced aligner (P2FA) is using to calculate log likelihood scores is more extreme in /tɹ/ and /dɹ/ clusters than in phonological affricates. The aligner is trained on data

from Supreme Court Justices, many of whom may have been older males producing fewer (if any) affricated variants before /ɹ/. It is certainly possible that younger speakers are producing affricated variants that derive higher log likelihood scores because the acoustic markers used to determine *probability of being an affricate* are more pronounced in these (affricate + [ɹ]) clusters, making them ‘more [tʃ]-like and [dʒ]-like than /tʃ/ and /dʒ/’ (to P2FA). It seems reasonable for a sound change to stop when it reaches a sound that is already present in the language, but it remains an open question whether affrication is still ongoing or has reached its conclusion.

### 2.3.1 Statistical Modeling

The *lmer* function in the *lme4* package (R Development Core Team, 2008) was used to generate mixed-effect linear regression models with A-Ratio as the dependent variable. A-Ratio was chosen over A-Score, because it provides a speaker-internal measure of /tɹ/ and /dɹ/ affrication within each speaker’s acoustic space (i.e., relative to productions of [t tʃ] and [d dʒ]). The models presented below include random intercepts for SPEAKER and WORD. The significance threshold is defined as a *t* value of greater than 2.0 or less than -2.0.

Models 1a and 1b<sup>2</sup> include a two-way interaction between BIRTHYEAR and SEX (a-ratio ~ birthyear \* sex + (1 | speaker) + (1 | word)), in both voiceless (Model 1a - TR) and voiced (Model 1b - DR) conditions. Table 2.3 shows a significant effect of birthyear vs. the reference level of A-Ratio (*t* value: 4.413) in voiceless condition, but no effect of sex vs. the reference level of A-Ratio (*t* value: 1.255), and no significant interaction between birthyear and sex (*t* value: -1.277). Table 2.4 shows a significant effect of birthyear vs. the reference level of A-Ratio (*t* value: 6.039) in the voiced condition, and also a significant effect of sex vs. the reference level of

---

<sup>2</sup>I considered using a random slopes structure for each word by speaker (1 + word|speaker), but the model would not converge, most likely due to the broad range of words in sociolinguistic interview data.

A-Ratio ( $t$  value: 2.315). However, there was also a significant interaction between birthyear and sex in Model 1b ( $t$  value: -2.348), with females showing higher A-Ratios than males, increasing at a faster rate over time. Although Model 1a shows no significant interaction between birthyear and sex, given that the  $t$  values are in the same direction as in Model 1b, it seems plausible that a similar change may be taking place in the voiceless condition, but to a lesser degree that is not detectable using this approach.

| AIC: 5840.6    | Estimate   | Std. Error | $t$ value |
|----------------|------------|------------|-----------|
| (Intercept)    | -21.916470 | 5.094587   | -4.302    |
| Birthyear      | 0.011479   | 0.002601   | 4.413     |
| SexM           | 9.293630   | 7.407287   | 1.255     |
| Birthyear:SexM | -0.004828  | 0.003780   | -1.277    |

**Table 2.3:** Summary of Model 1a – TR:  
 $a\text{-ratio} \sim \text{birthyear} * \text{sex} + (1 \mid \text{speaker}) + (1 \mid \text{word})$

| AIC: 2102.7    | Estimate   | Std. Error | $t$ value |
|----------------|------------|------------|-----------|
| (Intercept)    | -29.444961 | 4.985853   | -5.906    |
| Birthyear      | 0.015377   | 0.002546   | 6.039     |
| SexM           | 16.769154  | 7.243535   | 2.315     |
| Birthyear:SexM | -0.008678  | 0.003697   | -2.348    |

**Table 2.4:** Summary of Model 1b – DR:  
 $a\text{-ratio} \sim \text{birthyear} * \text{sex} + (1 \mid \text{speaker}) + (1 \mid \text{word})$

## 2.4 Phonologization or Contact? (Revisited)

The Figures (2.3 - 2.5) presented in the previous sections show what appears to be a gradual phonetic change over time in Raleigh, beginning somewhere around the middle of the twentieth century, suggesting phonologization of a phonetic effect. Generation 1 speakers are producing /tɹ/ and /dɹ/ sequences with comparatively

low A-Scores (more [t]-like and [d]-like), Generation 3 speakers are producing sequences with comparatively high A-Scores (more [tʃ]-like and [dʒ]-like), and Generation 2 speakers are producing sequences with more intermediate A-Scores. However, A-Score (and A-Ratio) values are plotted using speaker means, which has the potential to mask the introduction of new variants into the speech community through dialect contact. If, for example, following the installation of Research Triangle Park in Raleigh in the late 1950s, Northerners moving into the region imported with them affricated variants of /tɹ/ and /dɹ/ (more exaggerated versions of an already existing phonetic effect), and some Generation 2 Raleigh speakers began using a mix of existing and imported variants, we would see more intermediate mean A-Score (Figures 2.3 and 2.4) and A-Ratio (Figure 2.5) values for those speakers<sup>3</sup>.

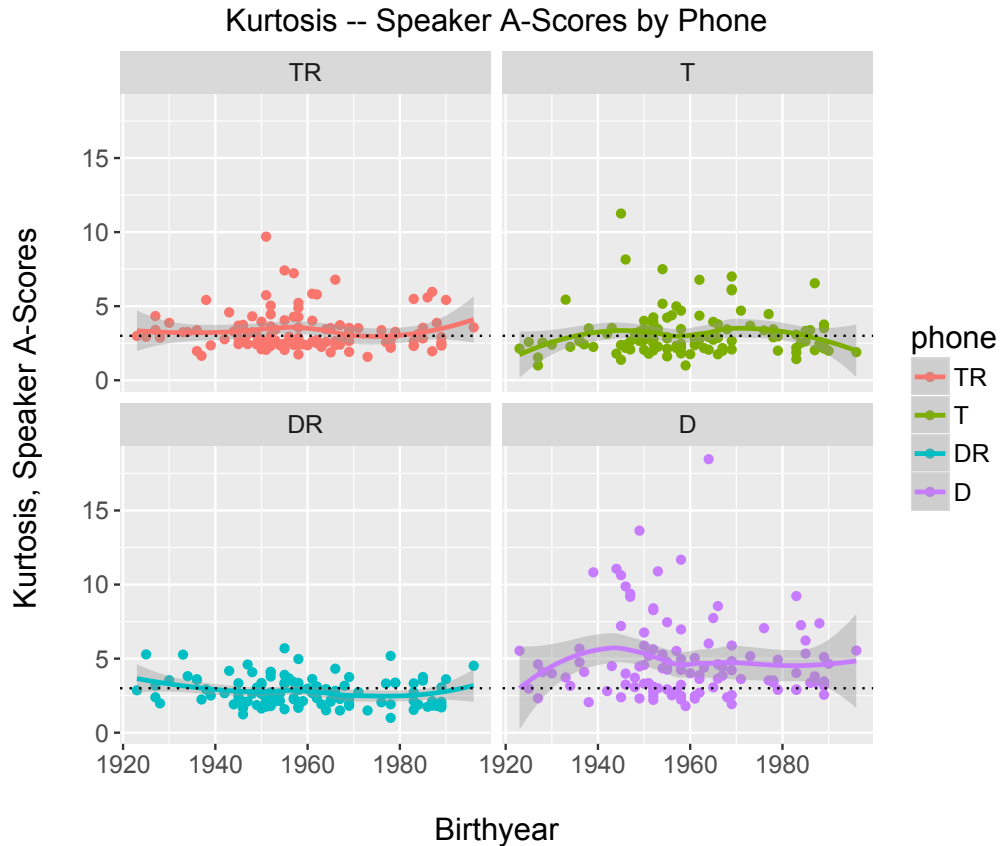
In a study on /ay/-raising in Philadelphia, Fruehwald (2016:384–386) examined the distribution of speaker data using kurtosis and standard deviation measures, in order to distinguish between continuous (phonologization) and categorical (dialect contact) variation. Kurtosis is a way of measuring bimodality, “where the lower the kurtosis, the greater the bimodality” (Darlington, 1970:19). If affricated variants of /tɹ/ and /dɹ/ were imported into Raleigh through dialect contact, we would expect to see a bimodal distribution (low kurtosis) for the Generation 2 speakers (those showing intermediate A-Score values) if both affricated and non-affricated (or less affricated) variants are present in their speech. In contrast, if this phenomenon is the result of a gradual phonetic change over time that has been phonologized, we would expect to see normal distribution within each speaker.

Modelled after Fruehwald (2016), Figure 2.6 plots the kurtosis of every speaker’s A-Score distribution by Birthyear for /t/ and /d/ tokens before /ɹ/ (left), and /t/ and /d/ tokens before vowels (right). The kurtosis of a normal distribution is

---

<sup>3</sup>Note that there is no particular reason to expect these Northerners to have had more affrication than Raleigh natives. A contact source for affrication is merely a possibility that can be explored in the Raleigh data.

represented by the horizontal dotted line at  $y = 3$ . The kurtosis of /t/ before /ɹ/ (TR, upper left) showed a normal distribution across birth years (mean: 3.3), and this same pattern was observed for the kurtosis of /d/ before /ɹ/ (DR, lower left, mean: 2.8). Following Fruehwald’s (2016) reasoning, if the change observed were the result of the introduction of a new variant into the Raleigh speech community, we would expect to see a dip in kurtosis at the mid-point of the change, where there is the greatest mix of both affricated and non-affricated variants. This is not the pattern we observe, however, as the trend in kurtosis across birth years is flat for both TR and DR.



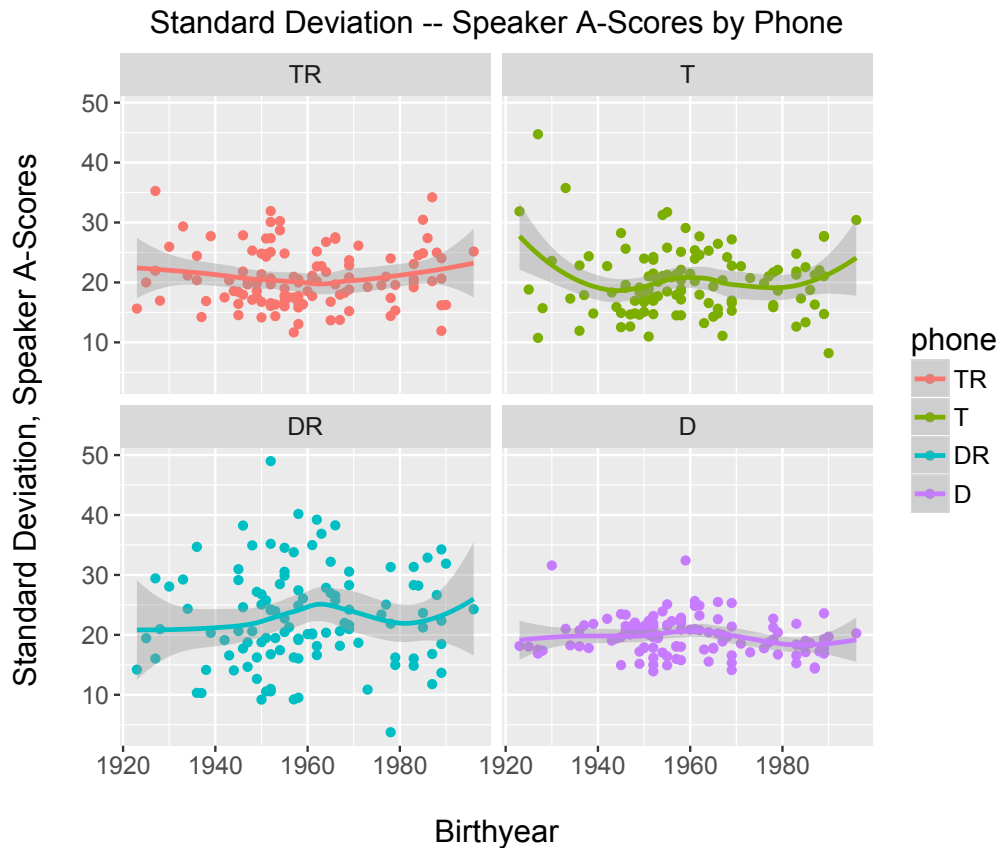
**Figure 2.6:** Estimated kurtosis of individual speakers’ A-Scores across Birthyear for /t/ and /d/ tokens before /ɹ/ (TR - upper left, DR - lower left), and before vowels (T - upper right, D - lower right). The kurtosis of a normal distribution is represented by the horizontal dotted line at  $y = 3$ . The  $y$ -scale is logarithmic.

Further, the kurtosis of /t/ before vowels (T, upper right) also showed a normal distribution across birth years (mean: 3.1), but the kurtosis of /d/ before vowels (D, lower right) was higher (mean: 4.9) though remained stable across birth years and may be a side-effect of the comparatively high number of data points for prevocalic /d/ (42% of the total words were of Word Type /d/ + vowel – see Table 2.2) and a greater number of outliers. If the sound change in progress in Raleigh were the result of categorical variation, due to dialect contact with Northerners using more affricated variants in their speech, we would expect to see noticeably different kurtosis patterns across pre-/ɪ/ (TR, DR) and prevocalic (T, D) categories, with prevocalic coronal stops showing flatter kurtosis over the course of the twentieth century (since a change in prevocalic stops is not underway), when compared with the patterns observed for /t/ and /d/ before /ɪ/. Since we do not, this provides further evidence to suggest that Raleigh speakers are not using a mix of affricated and non-affricated variants before /ɪ/.

Fruehwald (2016:385–386) also points out that two normal distributions (in our case, affricated and non-affricated variants of /tɪ/ and /dɪ/) will always have larger standard deviation than will a single distribution on its own. In other words, if the intermediate A-Score and A-Ratio values observed in Figures 2.3 - 2.5 are due to dialect contact and the resulting mixture of two variants at the height of the change, we would expect to see a peak in standard deviation near the middle of the twentieth century.

In Figure 2.7, which plots the standard deviation of individual speakers' A-Scores across Birthyear, we do not see patterns that might suggest the abrupt introduction of new affricated variants to the Raleigh speech community. As can be observed, the standard deviation of speakers' A-Scores for /t/ before /ɪ/ (TR, upper left, mean: 20.1) remains stable across birth years, with no peak at the midpoint of the change, and values that are in keeping with those observed for /t/ before vowels (T, upper

right, mean: 20.7). The standard deviation of speakers' A-Scores for /d/ before /ɪ/ (DR, lower left, mean: 22.8) also shows a relatively flat pattern across birth years, with no dramatic peak, but slightly more deviation as compared with /d/ before vowels (D, lower right, mean: 19.6). This modest increase in mean A-Score for DR (+3.2 over D, + 2.7 over TR) does not suggest normal distributions for two sets of variants, one with negative A-Score values, and one with positive A-Score values. Given that the overall range of A-Scores includes values between -100 and +100, we would expect to see much larger deviation. In addition, the kurtosis of /dɪ/ shows normal distribution for DR across birth years (Figure 2.6, lower left).



**Figure 2.7:** Standard deviation of individual speakers' A-Scores across Birthyear for /t/ and /d/ tokens before /ɪ/ (TR - upper left, DR - lower left), and before vowels (T - upper right, D - lower right).

Taken together, the kurtosis and standard deviation of A-Scores across birth

years do not point to the introduction of new variants to the speech community in Raleigh as a result of dialect contact, but rather to a gradual phonetic change over time that has been phonologized.

## 2.5 Discussion

This apparent time study into the phenomenon of /tɹ/ and /dɹ/ affrication provides strong evidence that a change was underway for English speakers in Raleigh, North Carolina near the middle of the twentieth century. The data show that older speakers born early in the twentieth century were producing /tɹ/ and /dɹ/ sequences that are more like [t] and [d], while younger speakers, born after the middle of the century, were producing these sequences more like [tʃ] and [dʒ] when examined using both A-Score and (speaker-normalized) A-Ratio measures. Females appear to be leading the change, as might be expected given that women have been shown to be early adopters of new variants (Labov, 2001). Further analysis of the data using kurtosis and standard deviation measures provides additional evidence to suggest that the change is not due to dialect contact, but rather to a gradual phonetic change over time.

Although these results are compelling, both A-Scores and A-Ratios are derived from log likelihoods assigned by the Penn Phonetics Lab Forced Aligner, using acoustic models that do not provide any transparent information about how the probability values are obtained. Further, the small number of speakers in certain birth years makes it difficult to assess, for example, how affricate-like younger females' productions of /tɹ/ and /dɹ/ are relative to their productions of prevocalic /tʃ/ and /dʒ/.

Results in this study provide strong motivation for a laboratory-based investigation into the phenomenon of /tɹ/ and /dɹ/ affrication. If a change is occurring in

Raleigh, what does the change look like in its current state, and what, if any, are the articulatory motivations for the change? The balance of the research presented in this dissertation explores this phenomenon using laboratory-based methods. Chapter 3 presents an articulatory study of speakers' productions of /tɹ/ and /dɹ/ sequences, and Chapter 4 presents a perceptual experiment designed to measure how speakers categorize affricated variants of /t/ before [ɹ].

# Chapter 3

## Production Study

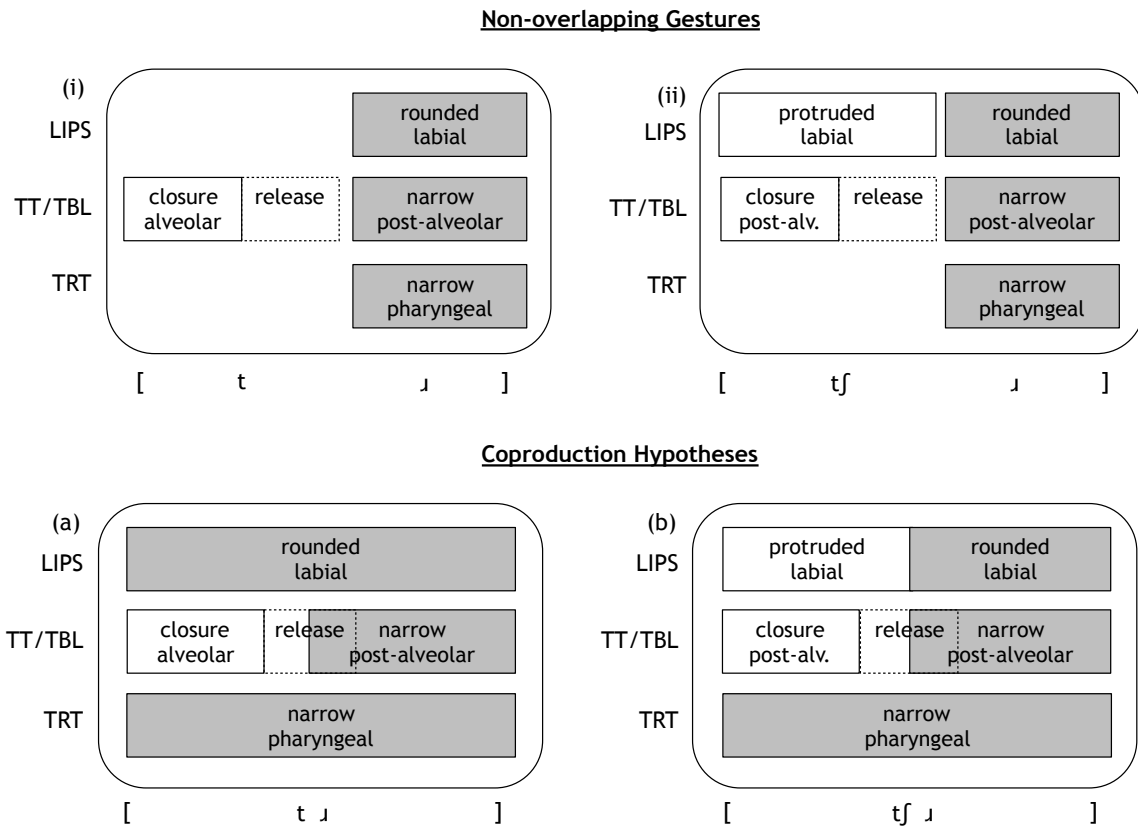
The Apparent Time Study results presented in Chapter 2 showed a sound change in progress in Raleigh, North Carolina. Speakers' /t/s and /d/s before /ɪ/ were increasingly more affricate-like over time, when examined using both A-Score and (speaker-normalized) A-Ratio measures. Based on these results, it was anticipated that participants in the Production Study would be more likely to produce variants of /t/ and /d/ before /ɪ/ that were affricated, given that all were born between 1985 and 1997, which aligns with the youngest (Generation 3) speaker birth years in the Apparent Time Study.

The Production Study was designed to examine the interaction between articulators during participants' productions of /t/ and /d/ before /ɪ/, and /t d tʃ dʒ/ before vowels. The goal of this stage of research was to answer the following question:

**Q2:** Are English /t/ and /d/ in /tɪ/ and /dɪ/ clusters articulated like prevocalic [t] and [d], like prevocalic [tʃ] and [dʒ], like neither, or like both?

This is an important question. Are participants producing [t]s and [d]s coarticulated with [ɪ] that, for aerodynamic reasons (Ohala and Solé, 2010), *sound* like affricates, or have participants adopted a new articulatory target before /ɪ/, suggesting phonologization of a coarticulatory effect?

As a starting point, we need to imagine what coarticulation of /t/ and /d/ before /ɹ/ would look like. Modelled after Browman and Goldstein (1989), Figure 3.1 presents hypothesized gestural scores for [tɹ] and [tʃɹ] clusters. (Similar hypothesized gestural scores are assumed for [dɹ] and [dʒɹ], differing only in terms of the timing of the onset of voicing.) The scores in the upper half of the figure (i, ii) represent these as *non-overlapping* sequences of gestures produced in slow succession<sup>1</sup>, while the scores in the lower half (a, b) represent hypotheses of these same gestures, *coproduced*.



**Figure 3.1:** Non-overlapping (upper, i-ii) and hypothesized coproduced (lower, a-b) gestural score diagrams for [tɹ] (left) and [tʃɹ] (right) sequences. The three articulatory tiers engaged in each score include the lips (LIPS), the tongue tip/blade (TT/TBL), and the tongue root (TRT).

<sup>1</sup>Non-overlapping gestures are included here for illustrative purposes only, to isolate the articulators for each gesture. They are not expected to occur in natural speech and instead represent an unrealistic scenario where there is no coarticulation. The focus of the discussions that follow will be on coproduction hypotheses a and b.

Figure 3.1 is a departure from Browman and Goldstein (1989) in several important ways<sup>2</sup>. First, the articulatory tier representing the tongue tip (TT) in Browman and Goldstein’s (1989) model has been broadened to include the tongue blade (TT/TBL). The tongue tip and tongue blade are grouped together here in a single articulatory tier because, although they involve slightly different regions of the tongue, the tip and blade do not act independently of one another (as we might see with, e.g. the tongue tip, and tongue dorsum or root). Second, a new articulatory tier has been introduced to represent engagement of the tongue root (TRT)<sup>3</sup>. Third, gestures involving complete oral closure (stops and affricates) have been elaborated to explicitly show the release following closure, which translates acoustically to aperiodic noise or voicelessness.

As illustrated in Figure 3.1, the only articulatory tier engaged during the production of [t] is TT/TBL (i, a), while both the LIPS and TT/TBL tiers are engaged during articulation of [tʃ] (ii, b). Production of [ɪ], a comparatively complex sound, engages articulators on all three tiers: TT/TBL, TRT, and (typically) LIPS (i, ii, a, b). The argument put forward here is that there may be nothing preventing early engagement of the lips and/or tongue root in anticipation of the following /ɪ/. In (a), the hypothesized gestural score for coproduced [tɪ], constriction at the lips (LIPS) and tongue root (TRT) for the /ɪ/ can begin as early as the onset of the closure for the stop (TT/TBL), as each articulator operates on a separate articulatory tier. In (b), the hypothesized gestural score for coproduced [tʃɪ], constriction at the tongue root for the /ɪ/ can also begin as early as the onset of the closure, whereas constriction at the lips for the /ɪ/ results in an articulatory conflict, since the LIPS tier is already engaged during production of /tʃ/. For both gestural scores (a, b), the

---

<sup>2</sup>Figure 3.1 makes no claims about Browman and Goldstein’s (1989) Articulatory Phonology model, and in no way attempts to refine it. Gestural scores are used here simply as a way of visually representing how gestures involving articulators on different tiers might be coproduced.

<sup>3</sup>Browman and Goldstein (1989:208) acknowledge “[o]ther oral tract variables that need to be implemented include an independent tongue root variable”.

narrow post-alveolar constriction of the tongue for articulation of /ɹ/ can also occur early, but only following release of the tongue after closure, as both gestures activate the same articulator on the same tier (TT/TBL). If either coproduction hypothesis is correct, our articulatory data should show lip and tongue root engagement during the closure interval for the stop/affricate gestures, and an [ɹ] tongue posture during the period of noise (aperiodicity) following release. It is anticipated that /ɹ/ tongue postures under either scenario will be predominantly bunched (tip-down), given that /ɹ/s produced following coronals are far more likely to be bunched, even for speakers who retroflex in other contexts (Mielke et al., 2016). In their study on s-retraction, Baker et al. (2011) found that all participants produced bunched /ɹ/ postures in the /stɹ/ context. Westbury et al. (1998:214) observed that speakers exhibited the least amount of variability in the /stɹ/ context, where /ɹ/ postures are more constrained. Mielke et al. (2016:114-115) have suggested “abandoning the bunched/retroflex distinction” for /ɹ/s next to lingual consonants, at least for some speakers, because in these contexts “the distinguishing features of bunched and/or retroflexed postures are both reduced, sometimes to the point of being indistinguishable”. In other words, while an interesting characteristic of North American English /ɹ/ is its variable behaviour, this variability is not likely to be a factor in the coarticulation of /tɹ/ and /dɹ/ sequences. It is expected that the articulatory complexity of the /ɹ/ gesture will be relevant here, and not its variability.

It is important to note that both coproduction hypotheses (a, b) could result in what *sounds* like an affricated /tɹ/: (b) begins with an affricate, and (a) involves a [t] released into a segment with a high tongue position, resulting in an extended period of turbulence that could be perceived as affrication (Ohala and Solé, 2010). In order to distinguish between these hypotheses, articulatory analysis will involve comparisons on each of the articulatory tiers beginning at closure. At the lips (LIPS), rounding at the mid-point of the closure interval would indicate early articulation

of the /ɹ/ gesture (a), whereas protrusion would suggest a gesture that is more [tʃ]-like than [ɹ]-like (or [t]-like). At the tongue tip/blade (TT/TBL), comparisons between /tɹ/, /tʃ/, and /t/ tongue postures will determine whether participants are producing more [t]-like (a) or [tʃ]-like (b) gestures. While retraction at the tongue root (TRT) will not distinguish between either of the hypotheses, evidence of retracted /tɹ/ and /dɹ/ tongue root postures at the mid-point of the closure interval will confirm the anticipated early articulation of the /ɹ/ in both scenarios.

## 3.1 Methodology

In order to test the coproduction hypotheses presented in Figure 3.1, ultrasound, video, and acoustic data of participants' productions of /tɹ/ /dɹ/ sequences were captured, along with their productions of prevocalic /t d tʃ dʒ/ sequences, in order to compare across tongue postures.

### 3.1.1 Participants

32 English speakers, recruited from the North Carolina State University community, participated in the Production Study. A subset<sup>4</sup> of these participants were included in the analysis that follows: 9 females and 3 males, ranging in age from 19 to 31 at the time of recording. Ten were born in North Carolina, one female subject was born in Wisconsin, and one male subject was born in Delaware. Participants were native speakers of English and reported no speech or hearing disorders, though S05 indicated “I had to learn how to say my Rs properly in elementary school” and S24 reported “a slight lisp sometimes”. Neither were considered to be problematic for the purposes of analysis – S05 produced adult-like /ɹ/s across contexts, and no lisp

---

<sup>4</sup>The twelve participants selected for the Production Study analysis were those whose acoustic, articulatory, and video data were aligned and ready when I was. For practical reasons, analysis of all 32 participants was outside the scope of this dissertation, due to the time required for manually processing that volume of articulatory data.

was detectable in S24’s speech.

| Participant | Sex    | Birth Year | Birth Place    |
|-------------|--------|------------|----------------|
| S03         | Female | 1996       | North Carolina |
| S05         | Female | 1995       | North Carolina |
| S06         | Female | 1996       | Wisconsin      |
| S09         | Male   | 1995       | Delaware       |
| S10         | Female | 1995       | North Carolina |
| S13         | Female | 1997       | North Carolina |
| S18         | Male   | 1995       | North Carolina |
| S19         | Female | 1996       | North Carolina |
| S23         | Male   | 1996       | North Carolina |
| S24         | Female | 1985       | North Carolina |
| S29         | Female | 1994       | North Carolina |
| S30         | Female | 1991       | North Carolina |

**Table 3.1:** Production Study participant details

### 3.1.2 Stimuli

Stimuli consisted of familiar monosyllabic English words (and one phonotactically probable nonword) beginning with: /tɪ t tʃ dɪ d dʒ/, across a range of unrounded vowel contexts that included high/low, front/back, and tense/lax distinctions: /i ɪ ʌ ɑ/. Proper names *Tron*<sup>5</sup> and *Jim* were used in two contexts, and a nonce word, *Dreep*, was used in one context where only one familiar lexical item could be found. Fillers consisted of words that were relevant to other studies being conducted at the NCSU Phonology Lab at the time of recording (see Appendix A for a complete list of experimental and filler tokens).

<sup>5</sup>Although the original *Tron*, a popular video game and film from the early 1980s, may not have been familiar to all participants, a more recent film, *Tron Legacy*, was released in 2010.

|          | /tɹ/  | /t/   | /tʃ/  | /dɹ/  | /d/  | /dʒ/  |
|----------|-------|-------|-------|-------|------|-------|
| <b>i</b> | tree  | tea   | cheap | dream | deem | jeans |
|          | treat | teach | cheek | Dreep | deep | jeep  |
| <b>ɪ</b> | trick | tick  | chick | drip  | dip  | Jim   |
|          | trip  | tip   | chip  | drift | dibs | jib   |
| <b>ʌ</b> | truck | tuck  | chuck | drug  | duck | jug   |
|          | trust | tub   | chum  | drum  | dumb | jump  |
| <b>ɑ</b> | trot  | talk  | chalk | drop  | dock | job   |
|          | Tron  | top   | chop  | draw  | dog  | jock  |

**Table 3.2:** Production stimuli

### 3.1.3 Procedures

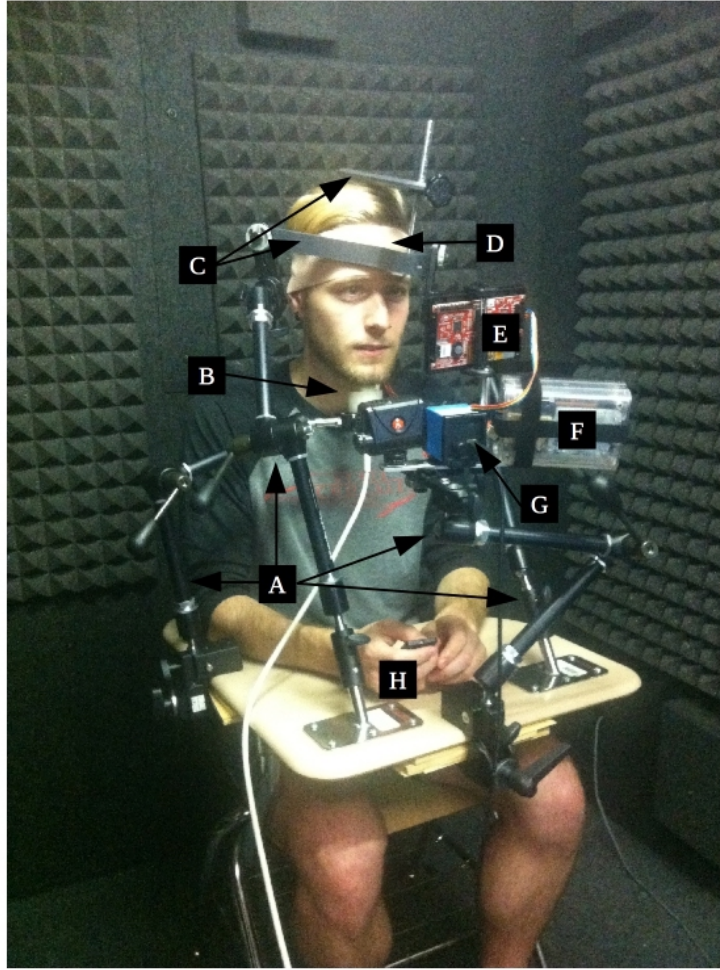
The Production Study was conducted in a sound-attenuated booth in the Phonology Lab at North Carolina State University (NCSU). The reading task, which included experimental phrases in the form “give me a X again”, took between 20 and 30 minutes to complete, following an initial set-up of approximately 10 minutes. Upon arrival, participants were asked to fill out a consent form (Appendix B.1) and were paid \$15 for their time. The Production experiment was immediately followed by a Perception experiment (Chapter 4) conducted in an adjacent sound-attenuated booth in the same lab. Following completion of both experiments, participants were asked to complete a language background questionnaire (Appendix B.2).

Ultrasound, video, and acoustic data were collected simultaneously using a Terason t3000 ultrasound machine, running Ultraspeech 1.3 (Hueber et al., 2008) in direct-to-disk mode. A microconvex array transducer (8MC3 3-8MHz, 90-degree field of view) was used to image the mid-sagittal plane of the tongue. Ultrasound images were captured in 640x480 pixel bitmaps, generated at a rate of 60 fps. Video images were also captured in 640x480 pixel bitmaps at 60 fps, using an Imaging Source DFK 21BU04 1/4" closed-circuit TV camera recording in grayscale mode. Audio was recorded at a sampling rate of 44,100 Hz with an Audio-Technica AT803 lavalier microphone, attached approximately one inch from the participant’s mouth

with an AT8418 instrument mounting clip. Audio was transmitted through a Sound-Devices USBPre 2.0 preamplifier.

### **3.1.3.1 Initial set-up**

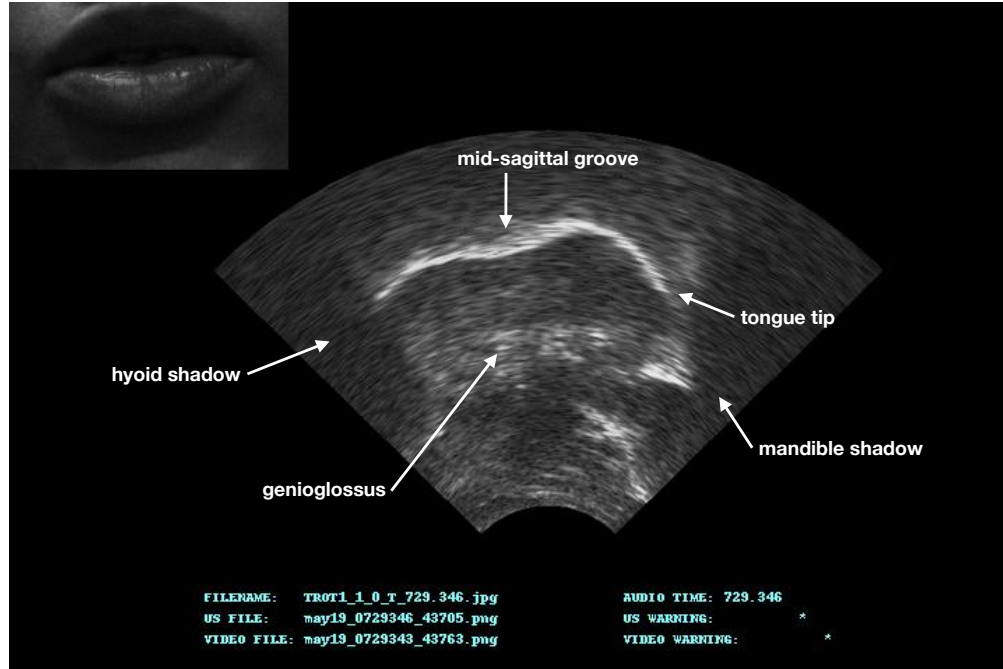
As illustrated in Figure 3.2, the participant was seated in a chair in the sound booth, and four locking adjustable Magic Arms (A) were used to position and stabilize the equipment during recording. The ultrasound transducer (B) was placed under the participant’s chin, and a headrest (C) was used to minimize vertical and horizontal movement and help stabilize the head in relation to the transducer. A headband (D) was used to hold head-orientation-tracking sensors in place, and an Arduino device (F) with a small screen (E) was programmed (Mielke, 2016) to provide feedback about head-orientation and movement (based on information communicated to the device from the sensors), and to display experimental phrases. A camera (G) was positioned to capture the participant’s lip movements during production of stimuli, and a remote control (H) for the Arduino device was used by the participant to advance through the experimental phrases displayed on the screen.



**Figure 3.2:** Participant set-up: (A) adjustable Magic Arms for positioning equipment; (B) ultrasound transducer; (C) headrest for head-stabilization; (D) headband to hold head-position sensors; (E) Arduino-compatible screen and (F) Arduino-based device for tracking head movement and presenting experimental phrases; (G) camera for recording lip movement; (H) remote control for advancing through stimuli.

During initial set-up, the participant was asked to find a comfortable position in the chair and fitted with the headband and sensors used to track head-orientation. The headrest was positioned in such a way as to provide the participant with head support, as well as tactile information about their vertical and horizontal position in space. At this stage, the participant was encouraged to ensure they were seated comfortably against the headrest, as they would be required to maintain a stable position throughout the experiment. The ultrasound transducer was positioned

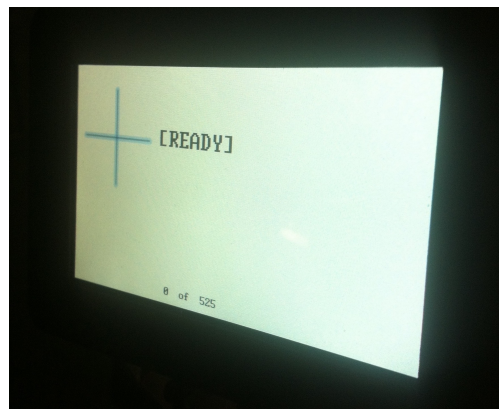
under the participant’s chin in order to capture mid-sagittal images of the tongue, and recording depth was adjusted for each participant to between 7 and 9 cm in order to image the full range of tongue movement within the oral cavity. Ultrasound and video image information was displayed on-screen (Figure 3.5), and was visible to the participant during set-up. Anatomical features visible in the ultrasound images were used to ensure proper transducer placement. The hyoid bone and mandible were used to assess front-to-back alignment, and the genioglossus muscle was used to assess side-to-side alignment (Figure 3.3), as these features can be seen in the ultrasound image if the transducer is properly aligned with the mid-sagittal plane. The participant was also asked to produce a sustained English [ə̃] – a sound that is typically produced with a deep mid-sagittal groove that can aid in locating the mid-sagittal plane. Once the transducer placement was established, the camera was adjusted and focused to capture a complete view of the participant’s lip and jaw movements during the recording. The microphone was then clipped approximately one inch from the participant’s mouth and the participant was prompted to produce some practice phrases that were of the same form as those in the experiment (e.g. “give me a unicorn again”), so that audio levels could be checked.



**Figure 3.3:** Anatomical features visible in ultrasound images used to ensure proper transducer placement along the mid-sagittal plane during recording.

A small Arduino-compatible screen was positioned at eye-level with the participant and at an appropriate viewing distance so as to be legible. Once it was confirmed that the participant was feeling confident that they would be able to maintain their position for the duration of the experiment, a baseline head position was obtained by synchronizing the head-band sensors with the Arduino device (through the push of a button on the remote control). Ensuring the participant maintained a consistent head position in relation to the ultrasound transducer was an important and necessary step in order to (later) be able to compare images across tokens during analysis (Stone, 2005; Davidson, 2006). As illustrated in Figure 3.4a, a blue cross and the message “READY” were displayed on the screen to indicate baseline head position. The participant was then instructed to move their head very slightly in order to see what would be displayed if they shifted from baseline. The message “please readjust” would appear on the screen (Figure 3.4b), along with yaw, pitch, and roll information about the participant’s head-orientation in relation to

baseline. The participant was then given the opportunity to practice returning to “READY” (baseline) position. It was explained to them that during the experiment they would be producing a series of phrases in the form “give me a X again”, and would be using the remote control to advance to the next phrase. Any head movement would result in the reappearance of the message “please readjust”, and they would be required to return to “READY” in order to advance to the next phrase. A total of 522 prompts were presented to each participant during the experiment, including 3 repetitions of 48 experimental stimuli (Table 3.2) and 3 repetitions of 126 fillers, with breaks at one third and two thirds of the way through the experiment. Stimuli were presented in one of four random orders.



(a) READY (baseline position)

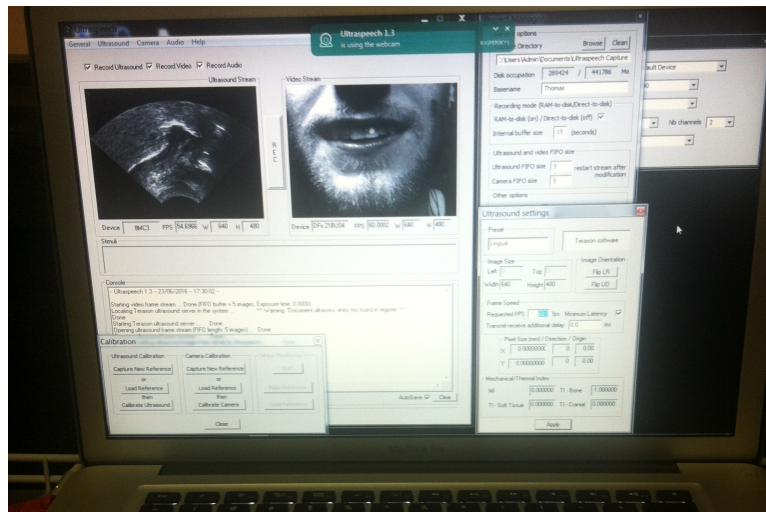


(b) please readjust (off baseline)

**Figure 3.4:** On-screen messages provide participants with information about their head position in relation to baseline.

### 3.1.3.2 Recording

At the start of recording and prior to producing any phrases, the participant was asked to sip water through a straw and hold it in their mouth in order to image the palate, then swallow the water. Next, the participant was asked to bite down on a tongue depressor and press their tongue up against it in order to image the occlusal plane. Once these two steps were completed, the remote control was handed to the participant, who was instructed to begin advancing through the word list *after* the experimenter had exited the booth and closed the door. The participant's productions were audible from outside of the booth through monitor speakers, which allowed the experimenter to identify whether the participant was moving through the list of tokens. If, at the beginning of the experiment, it became clear that a participant was having difficulty maintaining their baseline position (i.e. not advancing through the list of phrases at a reasonable pace), it was sometimes necessary for the experimenter to re-enter the sound booth and re-establish a baseline head position, which involved re-synchronizing the head-band sensors with the Arduino device.



**Figure 3.5:** On-screen ultrasound (left) and video (right) images, captured (with audio) using Ultraspeech 1.3 (Hueber et al., 2008). Here, a member of the NCSU Phonology Lab is producing a sustained English [æ], to align the transducer along the mid-sagittal plane.

### 3.1.4 Data Processing and Analysis

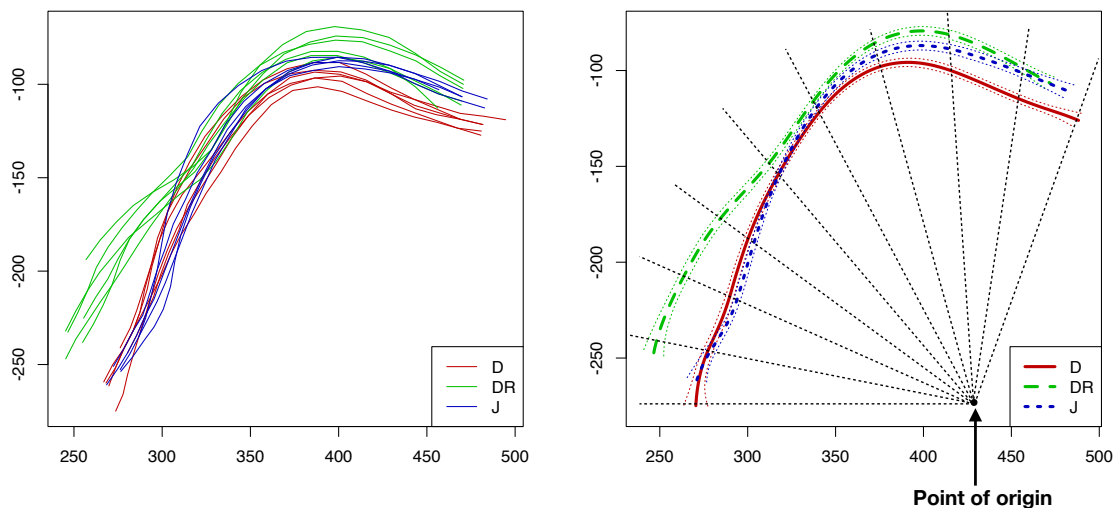
As described above, acoustic, video, and ultrasound data from 12 participants were collected, processed, and analyzed for the Production Study. Ultraspeech 1.3 (Hueber et al., 2008) was used to simultaneously capture acoustic, video, and ultrasound data. Automatic file-naming during recording provided synchronous time stamp information, which resulted in the straightforward alignment of all three streams of data (acoustic, video, and ultrasound).

#### 3.1.4.1 Acoustic analysis

Acoustic .wav files generated during production were segmented at the level of the phoneme using the Penn Phonetics Lab Forced Aligner (P2FA, Yuan and Liberman, 2008). P2FA was used to align each phrase independently, within the time interval when the corresponding text appeared on the screen. If the phrase was produced more than once by the participant, it was aligned with more than one repetition. All of the transcriptions were concatenated into one Praat (Boersma and Weenink, 2007) TextGrid. Word-initial /t d tʃ dʒ ɹ/ and following /ɹ/ and vowel intervals of experimental tokens were then hand corrected, as necessary, on a separate tier in Praat, and word-initial /t d tʃ dʒ/ were further segmented into consonant closure and aperiodic noise intervals. Only tokens where head orientation deviated by less than one degree along all three axes (i.e. where the participant's head orientation relative to the ultrasound transducer remained stable) were included at this and subsequent stages of analysis. Tokens that were mis-spoken (e.g. [tɹɹk] for *tuck*) were excluded from analysis. The segmentation was used to select ultrasound and video images at specific points of interest, and to calculate duration of the noise interval following release of the closure in /t d tʃ dʒ/ tokens. Duration was calculated by subtracting the time at the start of the aperiodicity from the time at the end of the aperiodicity, then multiplying this value by 1000 to convert to milliseconds.

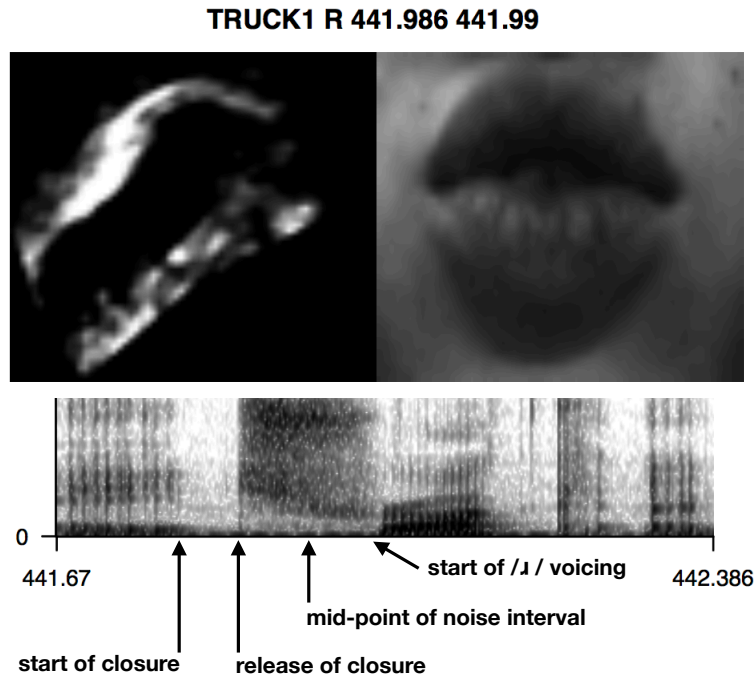
### 3.1.4.2 Articulatory analysis

Acoustic intervals segmented in Praat were used as reference points to select video and ultrasound images from the mid-point of the consonant closure and noise intervals, and from 20 milliseconds into the /ɹ/ and vowel intervals. A total of 4820 extracted ultrasound images were traced manually using Palatoglossotron (Baker, 2005). XY coordinates for each tongue trace (Figure 3.6, left), exported from Palatoglossatron, were transformed into polar coordinates, following the technique described by Mielke (2015). These polar coordinates were then used to generate Smoothing Spline ANOVAs (Figure 3.6, right) with an R script (R Development Core Team, 2008), developed by Mielke (2017). The Smoothing Spline ANOVA (SSANOVA) technique (Gu, 2002) is a way of determining whether there are statistically significant differences between smoothing splines (in our case, tongue curves), generated based on goodness of fit of the data being compared (Davidson, 2006). The SSANOVA method uses Bayesian confidence intervals (95%) to identify which regions of the curve (tongue contour) are statistically different from one another. For example, there may be significant differences between curves being compared at the tongue root, but not at the tongue tip/blade (e.g. Figure 3.6). In order to ensure curves were oriented similarly, tongue contours for each speaker were rotated so that their occlusal plane (a measure captured during recording) was horizontal. For analysis, and following Davidson (2006), tongue curves were divided into three regions (Figure 3.6, right) that roughly correspond to: the tongue tip/blade (upper right third), the tongue body/dorsum (middle third), and the tongue root (lower left third). The regions most relevant for examining participants' articulations of /t d tʃ dʒ r/ were the tongue tip/blade and the tongue root.



**Figure 3.6:** Left: raw tongue traces based on XY coordinates extracted from Palatoglossatron (Baker, 2005). Right: XY values transformed into polar coordinates (following Mielke, 2015), were used to generate SSANOVAs that were then divided into three analysis regions: the tongue tip/blade (upper right third), the tongue body/dorsum (middle third), and the tongue root (lower left third.)

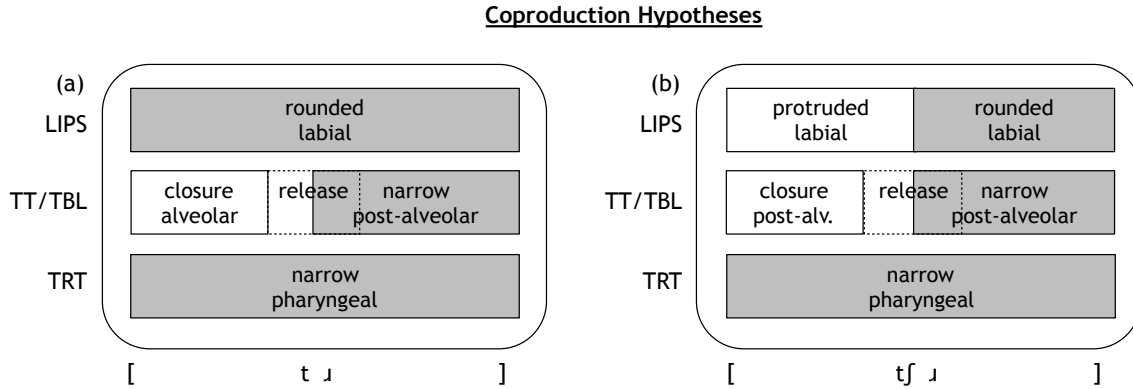
Analysis of the lips involved impressionistic coding of tokens in /ʌ/ vowel contexts, selected because it was reasoned that this lax, unrounded mid-vowel would have the least amount of influence on the lips. Lip postures from word-initial voiceless (*truck*, *tuck*, *chuck*) and voiced (*drug*, *duck*, *jug*) tokens were compared with one another and with word-initial /ɪ/ (*rub*) tokens. Image selection involved generating .pdf files of all the frames of selected tokens using a script developed by Mielke et al. (2017a), that takes matrices made from cropped, rotated, downsampled, and filtered versions of the original video, audio, and spectrogram images. As illustrated in Figure 3.7, using the spectrogram as a reference point, lip images were selected for comparison from the start of /t d tʃ dʒ/ at the beginning of the closure interval, from the release of /t d tʃ dʒ/ at the end of the closure interval, from the mid-point of the noise interval, and from the start of the /ɪ/ voicing in /tɪ/, /dɪ/, and word-initial /ɪ/ intervals.



**Figure 3.7:** Example of a single time point (start of /ɪ/ voicing in *truck*) across modalities (ultrasound, video, audio), generated using matrices from the original video and ultrasound images that were cropped, rotated, downsampled, and filtered.

## 3.2 Production Results

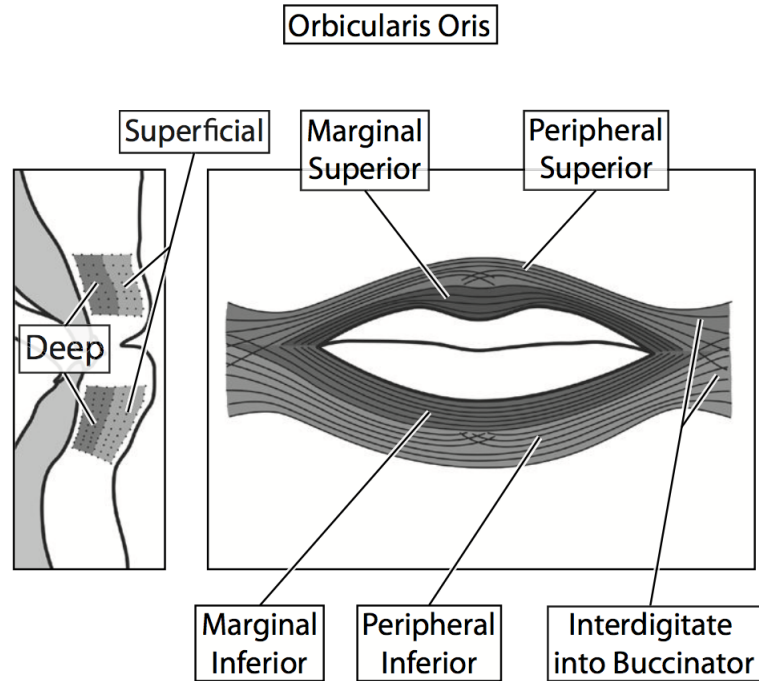
The results presented in this section include lip rounding data from video images, tongue contour data from ultrasound images, and consonant duration data from acoustic .wav files. As depicted in Figure 3.8 (a, b), it is anticipated that lip rounding and tongue root retraction may begin as early as the closure interval in /tɪ/ and /dɪ/ clusters, and articulation of the /ɪ/ tongue gesture may also begin early, during the period of voicelessness following release of the closure.



**Figure 3.8:** Gestural score diagrams for [tɪ] (left) and [tʃɪ] (right) sequences. The three articulatory tiers engaged in each score include the lips (LIPS), the tongue tip/blade (TT/TBL), and the tongue root (TRT). (First introduced in Figure 3.1, above, and presented again here for reference.)

### 3.2.1 Lip rounding

Although quantitative lip measures were outside the scope of this dissertation, a lot can be gleaned through visual inspection of lip images. By examining the shape of the lips and the engagement of the lip muscles, comparisons can be drawn between different articulations. The Orbicular Oris (OO), depicted in Figure 3.9 (Gick et al., 2013), is a muscle that comprises a substantial portion of the lips. The OO operates in a sphincter-like manner and is largely responsible for postures like rounding and protrusion. The marginal or inner part of the muscle (OOm), when contracted, results in a rounded, “pursed” shape with no lip protrusion and narrow lip aperture, while the peripheral or outer ring of the muscle (OOp), when contracted, forms a protruded constriction with wider aperture that may allow the inside of the lips to be partially visible.



**Figure 3.9:** Orbicular Oris (OO) muscle, responsible for lip rounding or pursing (OO marginal) and protrusion (OO peripheral). Image from Gick et al. (2013:193).

If hypothesis (a) in Figure 3.8 (above) is correct, we should see [ɪ]-like lip rounding as early as the closure interval for /t/. If hypothesis (b) is correct, we should see [tʃ]-like lip protrusion during the closure interval, and [ɪ]-like lip rounding during the noise interval following release.

Figure 3.10 shows video images of S03's lip postures during production of *tuck*, *truck*, *chuck*, and *rub*. This vowel context was selected for all the images that follow because /ʌ/ has the most neutral lip posture when compared with the other vowels in this study (/i ɪ ɑ/). The figure presents lip images from S03's productions of *tuck* (top row, a), *truck* (middle row, b), and *chuck* (bottom row, c), across three points in time: the start of the closure interval (i), the release of the closure interval (ii), and the mid-point of the noise interval following release (iii). Figure 3.10 also presents lip images from the start of the /ɪ/ in S03's productions of *truck* (middle

row, b) and *rub* (bottom row, d). Figure 3.11<sup>6</sup> shows images of S03’s lip postures during production of *duck* (a), *drug* (b), *jug* (c), and *rub* (d) at these same time points<sup>7</sup>.

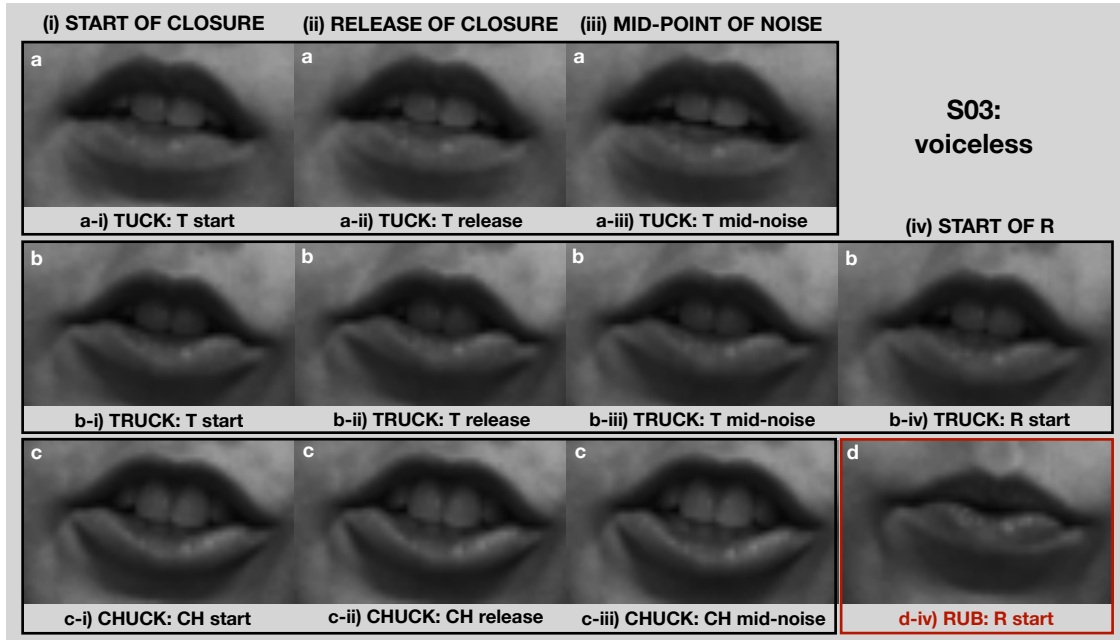
When comparing images within each figure (Figures 3.10 and 3.11), a number of observations can be made. For example, while the lips are open and the teeth visible during production of /t d/ in *tuck/duck* (a), /tɪ dɪ/ in *truck/drug* (b), and /tʃ dʒ/ in *chuck/jug* (c), the lips are pursed and the teeth are not visible during production of /ɪ/ in *rub* (d). This difference is most striking when comparing the start of /ɪ/ in *rub* (d-iv) and the start of /ɪ/ in *truck/drug* (b-iv), where in (d-iv) the OOm muscle is contracted, resulting in a more pursed, rounded posture with narrow lip aperture, whereas in (b-iv) the OOp muscle is contracted, resulting in a more protruded posture with wider lip aperture. Another noticeable difference can be observed at the corners of the mouth, in what Gick et al. (2013:197) refer to as “lateral compression, or pinching” – a mechanism used to brace the lips during constriction. In Figures 3.10 and 3.11, lateral compression at the corners of the mouth is evident in S03’s productions of *rub* (d), *truck/drug* (b), and *chuck/jug* (c), but not in *tuck/duck* (a), where the superior (upper) and inferior (lower) lips are not touching at any point along the two lips (no lateral compression). Comparing across postures at the start of closure, release of closure, and mid-point of the noise interval for /t d tʃ dʒ/, we see protruding lips at each time point in *truck/drug* (b) and *chuck/jug* (c), but not in *tuck/duck* (a), where there appears to be no bracing, protrusion, or rounding, suggesting little if any OO muscle engagement. Greater lip protrusion and wider aperture are observed in S03’s productions of /tʃ/ and /dʒ/ at the start (i) and the release (ii) of the closure interval, and the mid-point of the

---

<sup>6</sup>The image in Figure 3.11 (c) shows S03’s lip posture at the end rather than the mid-point of the noise interval, because the duration of the interval in voiced stops was too short to obtain a frame between release and end.

<sup>7</sup>Given the similarity between lip postures across voiceless and voiced tokens (e.g. *tuck*, *duck*) for all participants, and the shorter duration of the sibilant interval in voiced stops, only images from voiceless tokens will be shown from this point forward.

noise interval (iii) in *chuck/jug* (c), relative to productions of /t/ and /d/ at these same time points in *truck/drug* (b), where lip protrusion is evident, but lip aperture is smaller. Finally, we see increasing space between the superior and inferior teeth across time in *tuck/duck* (a), suggesting jaw lowering that is not present in any of the other articulations.

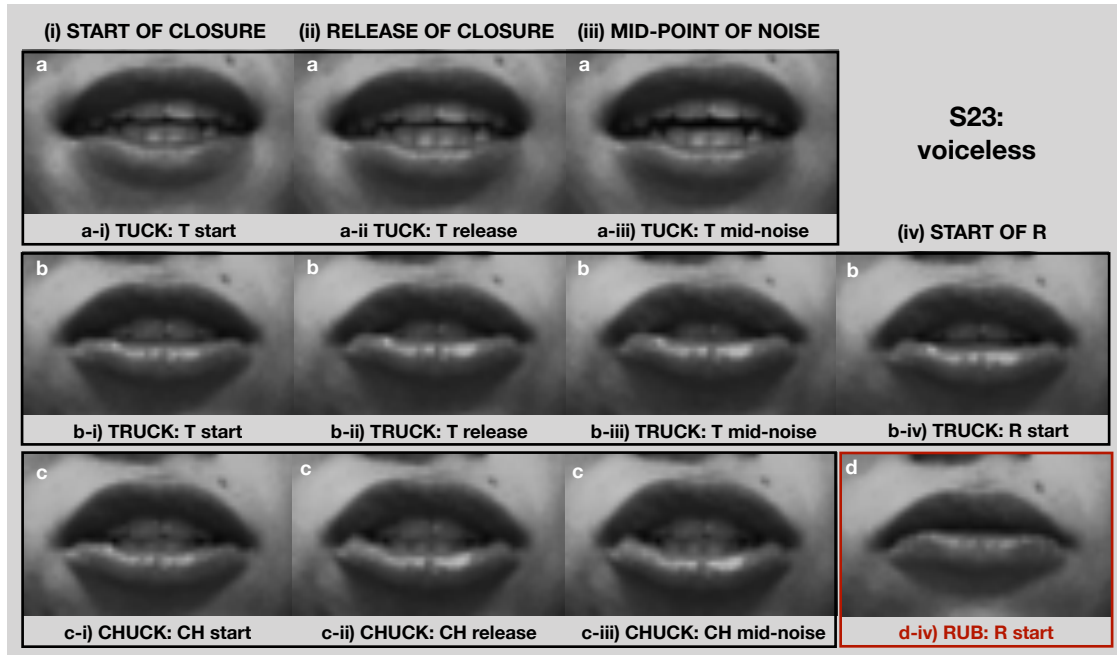


**Figure 3.10:** S03's lip postures for (a) *tuck*, (b) *truck*, (c) *chuck*, and (d) *rub*, at (i) the start of the closure interval, (ii) the release of the closure interval, (iii) the mid-point of the noise interval following release, and (iv) the start of the /ɹ/ (in *truck* and *rub*).



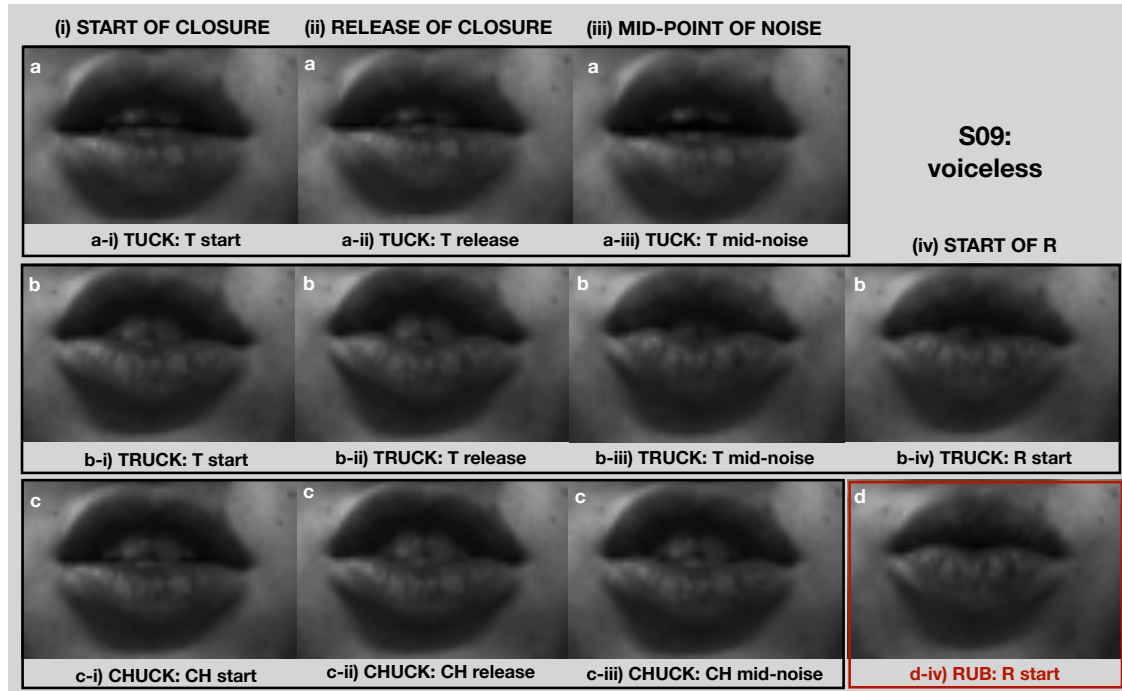
**Figure 3.11:** S03's lip postures for (a) *duck*, (b) *drug*, (c) *jug*, and (d) *rub*, at (i) the start of the closure interval, (ii) the release of the closure interval, (iii) the mid-point of the noise interval, and (iv) the start of the /ɹ/ (in *drug* and *rub*).

The general lip rounding pattern described for S03 holds across participants, though there is some inter-speaker variation in terms of the amount of lip aperture, protrusion, presence of lateral compression, and spacing between superior and inferior teeth. In Figure 3.12, for example, S23 shows a less rounded, more spread, lip posture than S03, but exhibits similar lip protrusion and lateral compression in *truck* (b) and *chuck* (c), that is not present in *tuck* (a). Similar to S03, S23 also shows spacing between the superior and inferior teeth during production of *tuck* (a), suggesting jaw lowering not present in the other articulations (b-c), but for S23 this spacing exhibits more of an increase (more active jaw-lowering) by the mid-point of the noise interval (iii).



**Figure 3.12:** S23's lip postures for (a) *tuck*, (b) *truck*, (c) *chuck*, and (d) *rub*, at (i) the start of the closure interval, (ii) the release of the closure interval, (iii) the mid-point of the noise interval, and (iv) the start of the /ɹ/ (in *truck* and *rub*).

S09 (Figure 3.13) generally shows less variation in lip aperture, as well as an increase in rounding over the course of *truck* (b-i to b-iv), such that the start of /ɹ/ in *truck* (b-iv) appears more similar to the start of /ɹ/ in *rub* (d-iv), when compared with other participants' lip postures at these time points. Despite this apparent similarity, however, the inside of the lips (and teeth) are partially visible in *truck* (b-iv), suggesting OOp muscle activation, whereas *rub* (d-iv) shows a more pursed posture with lips rounded inward, suggesting OOm muscle activation.

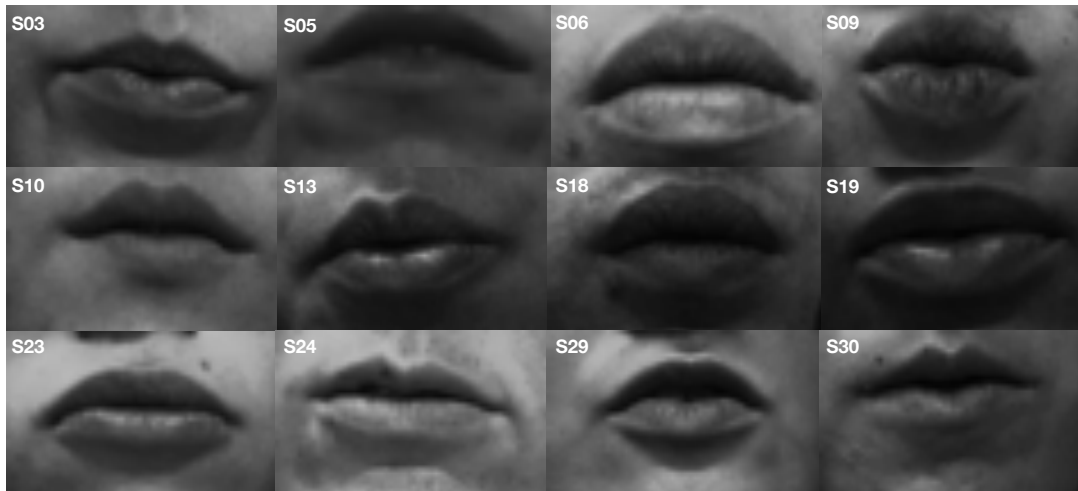


**Figure 3.13:** S09's lip postures for (a) *tuck*, (b) *truck*, (c) *chuck*, and (d) *rub*, at (i) the start of the closure interval, (ii) the release of the closure interval, (iii) the mid-point of the noise interval, and (iv) the start of the /ɹ/ (in *truck* and *rub*).

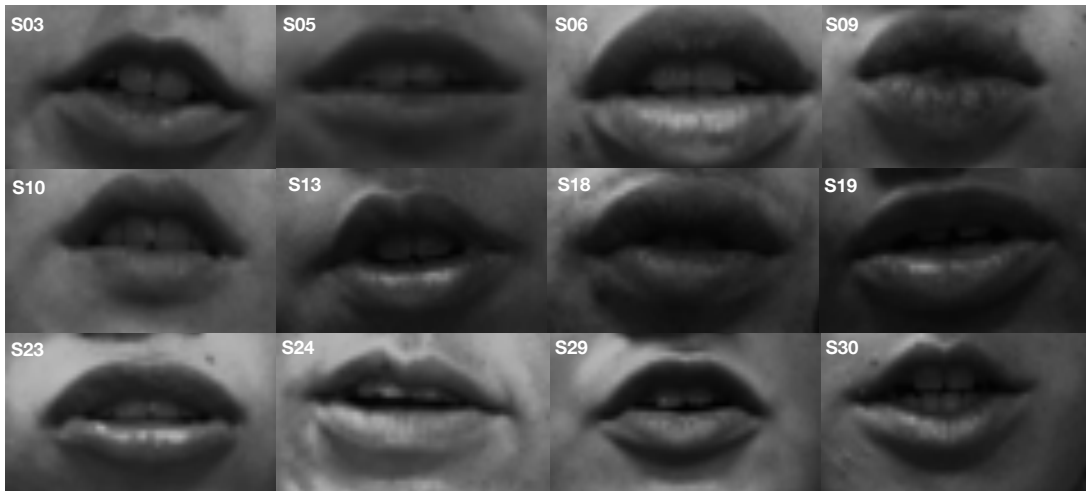
The figures that follow present lip images for all participants at the start of the /ɹ/ in *rub* and *truck*, clearly illustrating that the two /ɹ/s are not the same. Participants show pursed lip postures, no protrusion, and narrow lip aperture at the start of the /ɹ/ in *rub* (Figure 3.14), as compared with their more protruded lip postures and relatively wider lip aperture at the start of the /ɹ/ in *truck* (Figure 3.15), suggesting engagement of different regions of the Orbicular Oris muscle (OOm for *rub* postures, OOp for *truck* postures), and different articulatory strategies for producing these two /ɹ/s.

Overall, the results presented here align most closely with hypothesis (b) in Figure 3.1: participants show early engagement of the LIPS articulatory tier, with lip protrusion evident during the closure interval. However, this protrusion persists through the noise interval and start of the /ɹ/ gesture, which is not consistent with

either hypothesis.



**Figure 3.14:** Start of /ɪ/ in *rub* for all participants, showing contracted OOm with pursed lip postures, no protrusion, and narrow lip aperture, as compared with the start of /ɪ/ in *truck* (Figure 3.15).



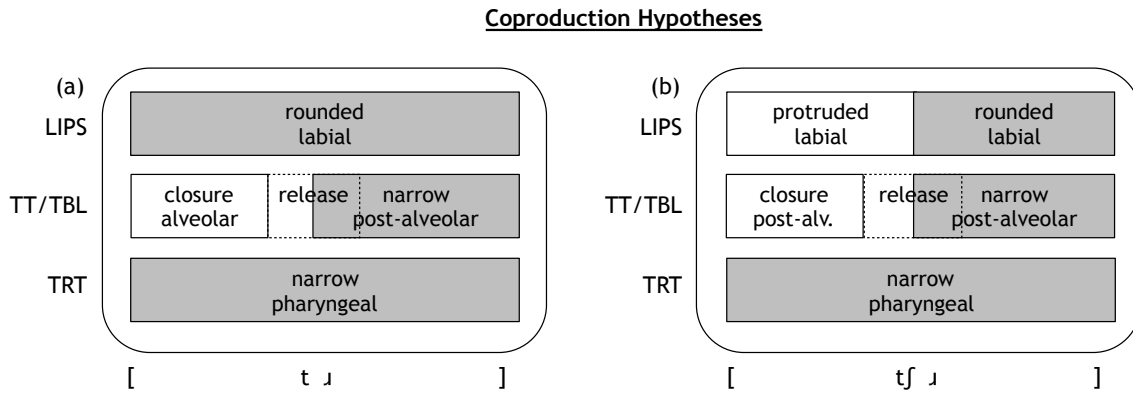
**Figure 3.15:** Start of /ɪ/ in *truck* for all participants, showing contracted OOp with protruded lip postures and relatively wider lip aperture, as compared with the start of /ɪ/ in *rub* (Figure 3.14).

### 3.2.2 Tongue contours

Participants were producing lip postures that most closely resembled [tʃ] and [dʒ], and (crucially) not [t], [d], or [ɹ], as early as the mid-point of the closure interval. If tongue posture data are consistent with lip posture data, we might expect to see evidence of tongue tip/blade contours for /t/ and /d/ before /ɹ/ that are similar to those for /tʃ/ and /dʒ/ before vowels at the same time point. Regardless of posture at the tongue tip/blade, we also anticipate evidence of early tongue root retraction in /tɹ/ and /dɹ/ sequences.

As illustrated in Figure 3.16, both hypotheses (a) and (b) predict that participants will exhibit tongue root retraction (TRT tier) as early as the closure interval. Where (a) and (b) differ is at the tongue tip/blade (TT/TBL tier). If hypothesis (a) is correct, participants' tongue contours for /t/ and /d/ before /ɹ/ should be similar to their /t/ and /d/ contours before vowels at the mid-point of the closure interval because they are not producing distinct articulatory targets for /t/ and /d/ before /ɹ/. If hypothesis (b) is more closely aligned with the data, in which participants *are* producing distinct articulatory targets for /t/ and /d/ before /ɹ/ (that are different from their articulatory targets for /t/ and /d/ before vowels), tongue contours in /tɹ/ and /dɹ/ sequences may resemble prevocalic /tʃ/ and /dʒ/ contours as early as the mid-point of closure. By the mid-point of the noise interval following release, hypotheses (a) and (b) predict more [ɹ]-like tongue tip/blade postures (TT/TBL tier). Given that [ʃ] and (tip-down/bunched) [ɹ] both involve tongue blade raising with post-alveolar constriction, it is plausible that participants' articulations of /tɹ/ and /dɹ/ may look articulatorily similar to /tʃ/ and /dʒ/ under either scenario. One possible way to disambiguate [tʃ] or [dʒ] from [ɹ] is to look for concavity at the tongue dorsum, an articulatory characteristic of English bunched [ɹ]. While tongue root retraction is also an articulatory characteristic of an English [ɹ], it may not help to distinguish [tʃ] or [dʒ] from [ɹ] because (as hypothesized), tongue root retraction

may co-occur with the stop/affricate as early as the closure interval under either scenario, since neither articulation engages the tongue root tier.



**Figure 3.16:** Gestural score diagrams for [tɪ] (left) and [tʃɪ] (right) sequences. The three articulatory tiers engaged in each score include the lips (LIPS), the tongue tip/blade (TT/TBL), and the tongue root (TRT). (First introduced in Figure 3.1, above, and presented again here for reference.)

### 3.2.2.1 Tongue contour analysis

The SSANOVAs presented here show statistical comparisons of tongue contours for /t/ and /d/ before /ɪ/ (TR, DR), and /t/ and /d/ (T, D) and /tʃ/ and /dʒ/ (CH, J) before vowels, at two points in time: the mid-point of the closure interval, and the mid-point of the noise interval following release. Non-overlapping confidence intervals along each contour indicate statistically significant differences between the curves. As described above, for the purposes of analysis, each tongue contour is divided into three sections, roughly corresponding to the tongue tip/blade (upper right third), tongue body/dorsum (middle third), and tongue root (lower left third) (following Davidson, 2006). The focus of this articulatory investigation is on the tongue tip/blade and the tongue root, as these are the regions where it is hypothesized that /t/ and /d/ before /ɪ/ will pattern differently relative to /t/ and /d/ before vowels. The figures that follow show participants' typical and atypical behaviour patterns across the two time points (closure and noise), in each of the four

(front /i ɪ/ and back /ʌ ɑ/) vowel contexts, with comparisons between contours at the tongue root and the tongue tip/blade.

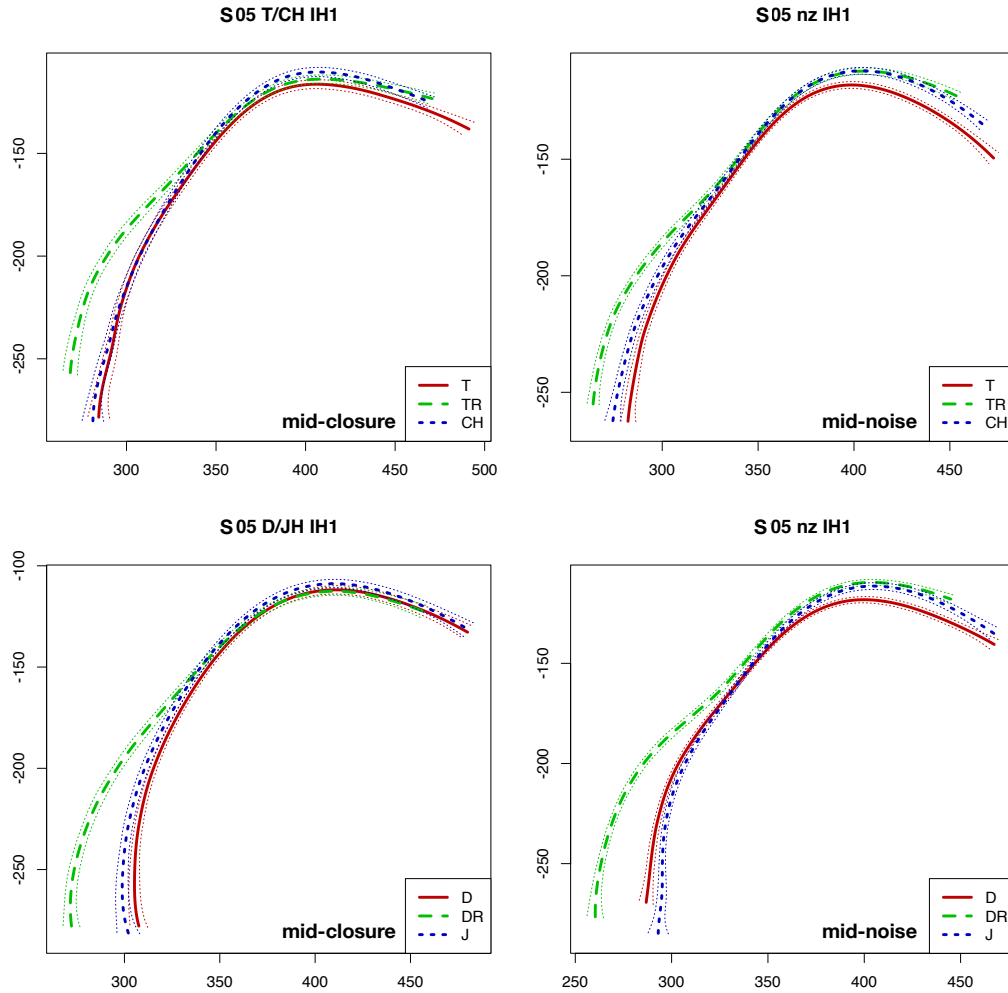
### 3.2.2.2 Front vowel contexts

In front vowel contexts (e.g., /t.ɪt d.ɪm/, /t.ɪp d.ɪp/), participants typically<sup>8</sup> showed retracted tongue roots in /tɹ/ and /dɹ/ contours at both the mid-point of the closure interval and the mid-point of the noise interval, when compared with /t/ and /d/ and /tʃ/ and /dʒ/ contours, which were relatively more advanced (TR>T,CH, DR>D,J). At the tongue tip/blade, participants exhibited no significant differences between contours (TR=CH=T, DR=J=D) mid-closure in both front vowel contexts, but while the same pattern was observed at the mid-point of the noise interval in /i/ vowel contexts, in /ɪ/ vowel contexts, participants' /tɹ/ and /tʃ/, and /dɹ/ and /dʒ/ contours typically patterned together, displaying significantly higher curves at the tongue tip/blade, relative to prevocalic /t/ and /d/ (TR,CH>T, DR,J>D).

Figure 3.17 shows S05 exhibiting typical voiceless (upper) and voiced (lower) tongue contours in /ɪ/ vowel contexts, at the mid-point of the closure interval (left), and the mid-point of the noise interval (right). As illustrated, tongue roots (lower left third of each SSANOVA) in voiceless /tɹ/ and voiced /dɹ/ contours (TR, DR, dashed green lines) are significantly farther back (from the point of origin, see Figure 3.6) at both time points, relative to prevocalic /t/ and /d/ (T, D, solid red lines) and /tʃ/ and /dʒ/ (CH, J, dotted blue lines) contours (TR>T,CH, DR>D,J). At the tongue tip/blade (upper right third of each SSANOVA), S05 showed no significant differences between contours (TR=T=CH, DR=D=J) at the mid-point of the closure interval (left), but at the mid-point of the noise interval (right), showed /tɹ/ and /tʃ/ (upper), and /dɹ/ and /dʒ/ (lower) contours patterning together, with higher tongue postures relative to /t/ and /d/ contours (TR,CH>T, DR,J>D).

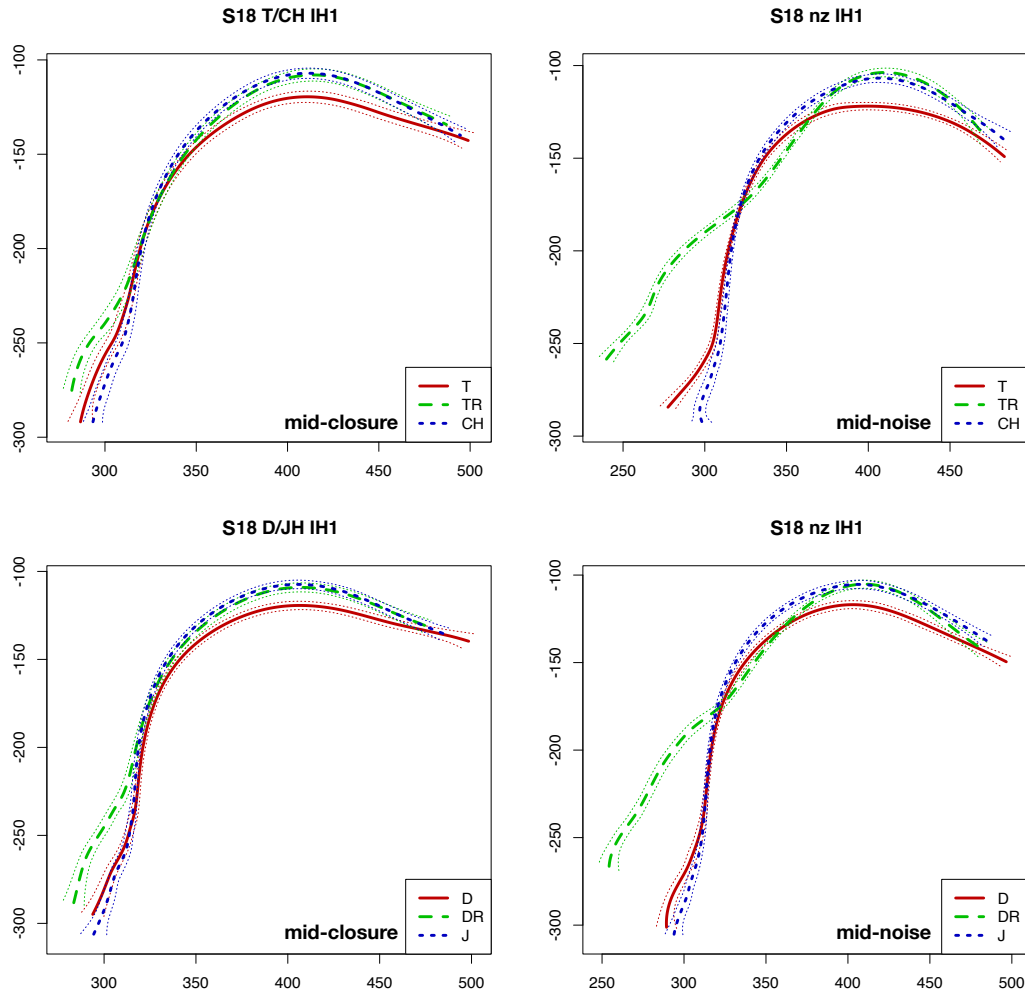
---

<sup>8</sup>Summary tables with participant totals for all typical and atypical patterns at the tongue root (Table 3.3) and tongue tip/blade (Table 3.4) are presented later in the discussion.



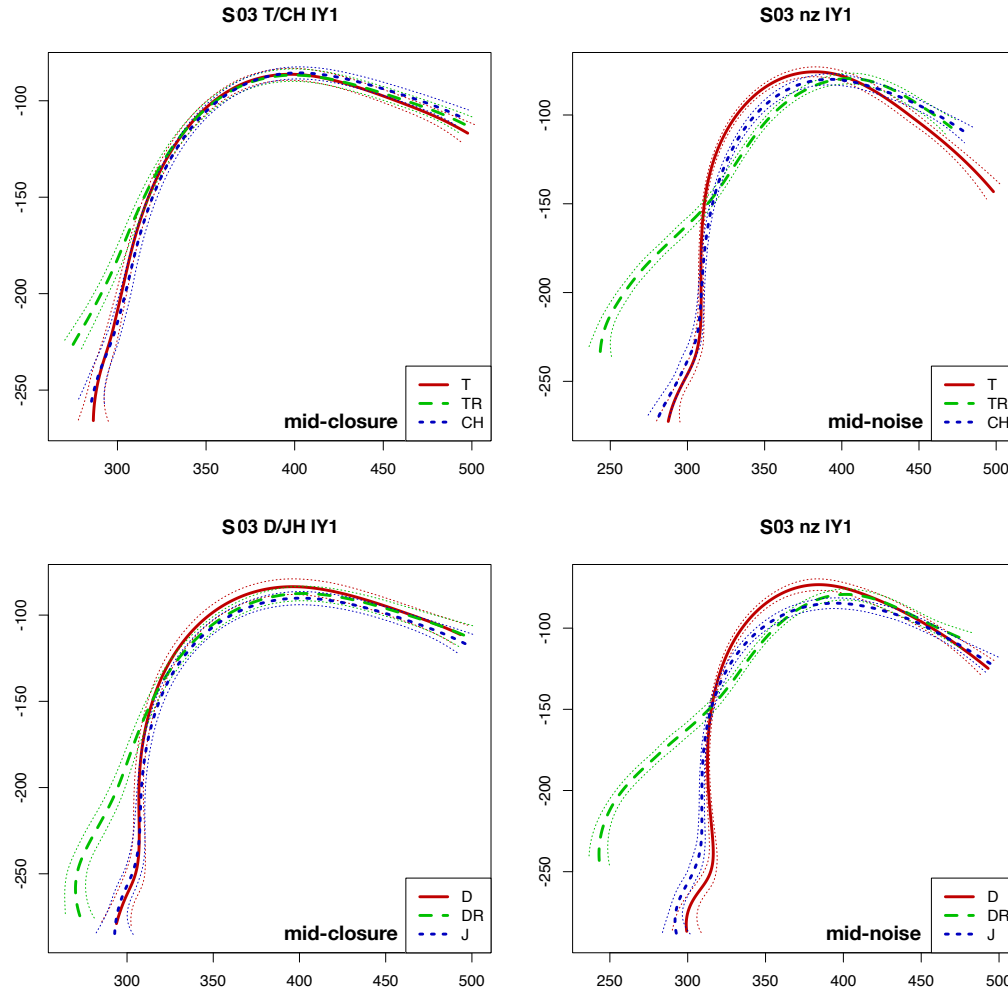
**Figure 3.17:** S05: typical patterns in /ɪ/ vowel contexts at the tongue root and tongue tip/blade in voiceless (upper, T TR CH) and voiced (lower, D DR J) contours, at the mid-point of the closure interval (left), and the mid-point of the noise interval (right).

Several participants deviated from the group in /ɪ/ vowel contexts at the tongue tip/blade (S18, Figure 3.18, left), showing /tɪ/ and /tʃ/ (upper), and /dɪ/ and /dʒ/ (lower) contours patterning together relative to /t/ and /d/ (TR,CH>T, DR,J>D) at the mid-point of the closure interval (i.e., earlier than the rest of the group), but exhibiting typical behaviour elsewhere.



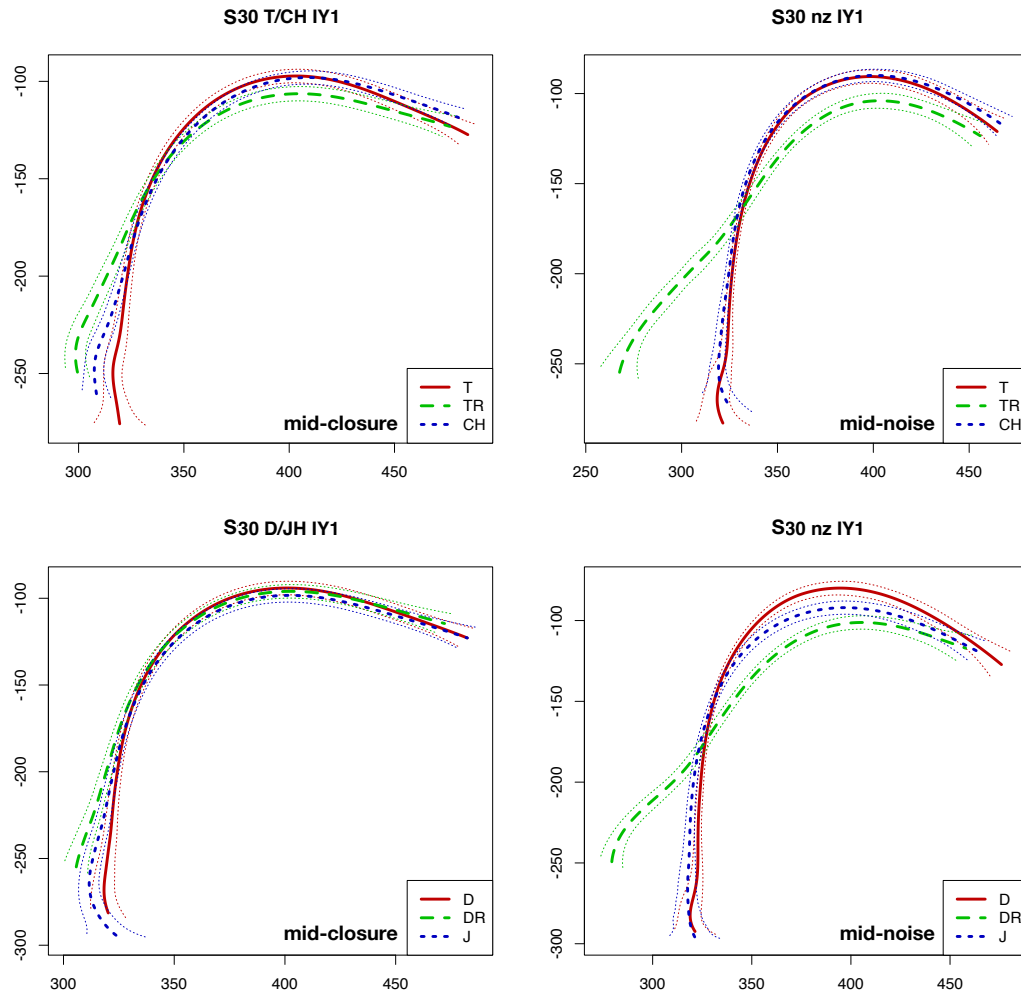
**Figure 3.18:** S18: example of participant deviating from the group in /ɪ/ vowel contexts at the mid-point of the closure interval (left) in voiceless /tɪ/ (upper, TR) and voiced /dɪ/ (lower, DR) contours at the tongue tip/blade, but patterning with the group elsewhere.

In /i/ vowel contexts, participants showed similar tongue root patterns as in /ɪ/ vowel contexts, with retracted tongue roots in /tɪ/ and /dɪ/ at both time points, relative to the comparatively more advanced tongue roots in /t tʃ d dʒ/ (TR>T,CH, DR>D,J), but typically exhibited no significant differences between contours (TR=T=CH, DR=D=J) at the tongue tip/blade at either the mid-point of the closure interval or the mid-point of the noise interval (S03, Figure 3.19).



**Figure 3.19:** S03: typical /tɪ/ (upper, TR) and /dɪ/ (lower, DR) patterns in /i/ vowel contexts at the tongue root and tongue tip/blade, at the mid-point of the closure interval (left), and the mid-point of the noise interval (right).

Several participants deviated from the typical group pattern in /i/ vowel contexts at the tongue tip/blade, showing comparatively (though not always significantly) lower tongue tip/blade contours in /tɪ/ and/or /dɪ/ sequences relative to the other contours (CH,T>TR, J,D>DR), particularly at the mid-point of the noise interval (S30, Figure 3.20).

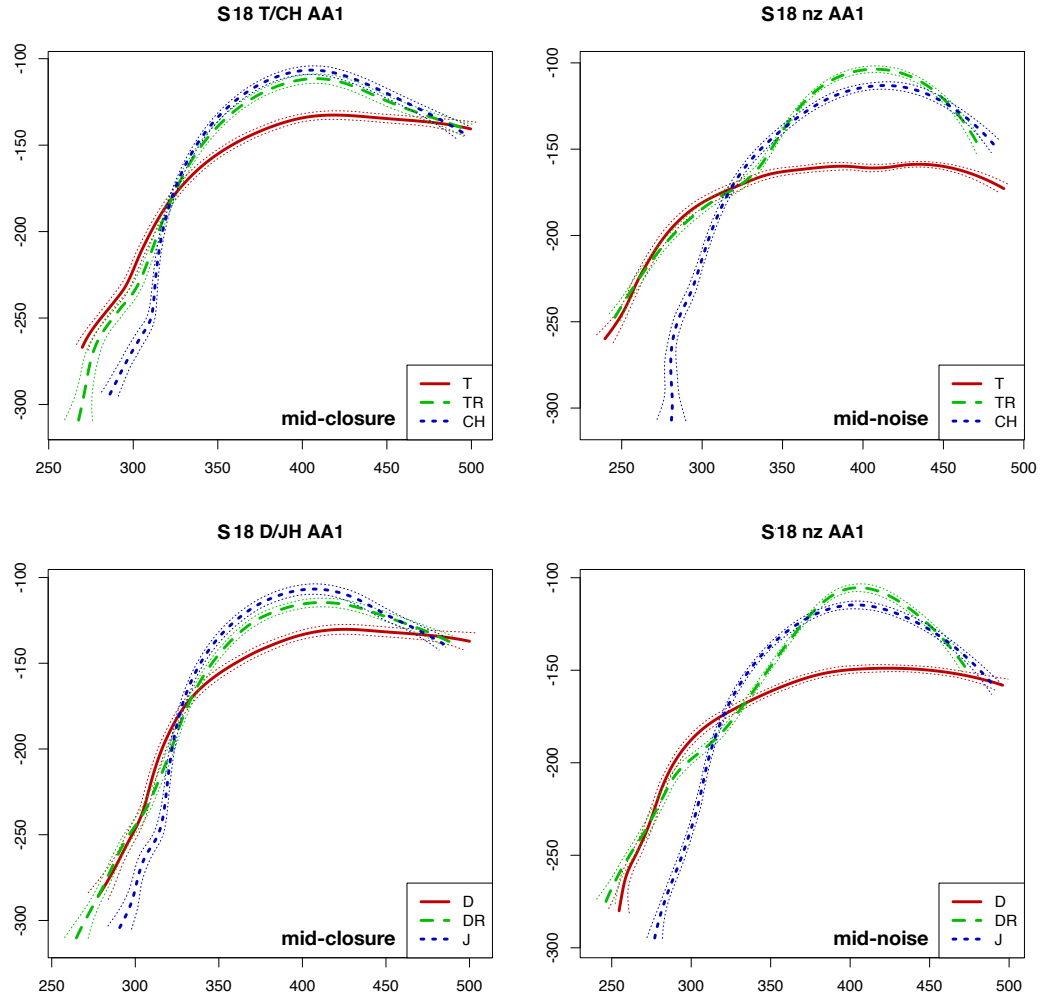


**Figure 3.20:** S30: deviating from the group in /i/ vowel contexts, exhibiting lower tongue tip/blade contours in /tɪ/ (upper, TR) or /dɪ/ (lower, DR) sequences relative to the other contours, but otherwise patterning with the group.

### 3.2.2.3 Back vowel contexts

In back vowel contexts (e.g., /tɪɫk dɪɫg/, /tɪɫt dɪɫp/), more variation was observed at both time points. Participants typically exhibited one of three patterns at the tongue root, showing retracted /tɪ/ and /dɪ/ contours relative to /t tʃ/ and /d dʒ/ (TR>T,CH, DR>D,J), retracted /tɪ/ and /t/, and /dɪ/ and /d/ contours relative to /tʃ/ and /dʒ/ (TR,T>CH, DR,D>J), or no significant differences between contours (TR=T=CH, DR=D=J). At the tongue tip/blade, participants typically showed /tɪ/ and /tʃ/, and /dɪ/ and /dʒ/ contours patterning together, with significantly higher postures relative to /t/ and /d/ (TR,CH>T, DR,J>D) at both time points, though several participants showed no significant differences between tongue tip/blade contours (TR=T=CH, DR=D=J) at the mid-point of the closure interval, particularly in /ɑ/ vowel contexts. One participant deviated from these tongue tip/blade patterns mid-closure, exhibiting /tɪ/ and /t/, and /dɪ/ and /d/ contours patterning together, with significantly lower tongue tip/blade postures than /tʃ/ and /dʒ/ (CH>TR,T, J>DR,D). Additionally, at the mid-point of the noise interval, several participants exhibited three distinct tongue contours at the tongue tip/blade, with /tɪ/ and /dɪ/ showing higher postures than /tʃ/ and /dʒ/, and /tʃ/ and /dʒ/ showing higher postures than /t/ and /d/ (TR>CH>T, DR>J>D).

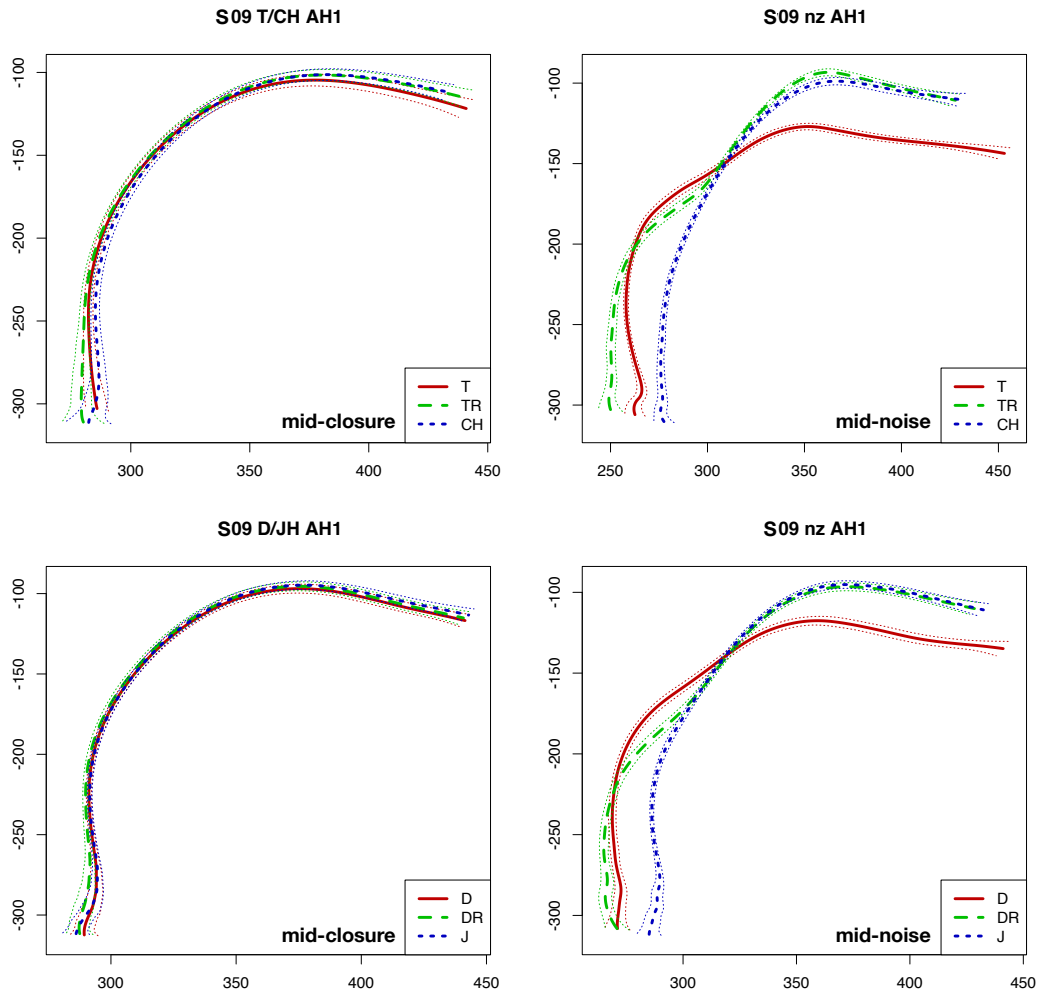
Figure 3.21 presents typical back vowel patterns in /ɑ/ vowel contexts. Here, S18 shows /tɪ/ and /t/, and /dɪ/ and /d/ patterning together at the tongue root, at both the mid-point of the closure interval (left), and the mid-point of the noise interval (right), exhibiting more retracted postures relative to /tʃ/ and /dʒ/ (TR,T>CH, DR,D>J). In contrast, at the tongue tip/blade, S18 shows /tɪ/ and /tʃ/, and /dɪ/ and /dʒ/ patterning together at the two time points, with significantly higher postures relative to /t/ and /d/ (TR,CH>T, DR,J>D).



**Figure 3.21:** S18: typical tongue contour patterns in /a/ vowel contexts at the tongue root and tongue tip/blade in voiceless (upper, T TR CH) and voiced (lower, D DR J) contours, at the mid-point of the closure interval (left), and the mid-point of the noise interval (right).

Figure 3.22 illustrates typical back vowel patterns in /ʌ/ vowel contexts. At the mid-point of the closure interval (left), S09 exhibits no significant differences between contours at the tongue root or at the tongue tip/blade (TR=T=CH, DR=D=J). At the mid-point of the noise interval (right), S09 shows a more retracted /tɪ/ tongue root posture relative to /t/ and /tʃ/ (TR>T,CH), and more retracted /dɪ/ and /d/ tongue root postures relative to /dʒ/ (DR,D>J), while at the tongue tip/blade, shows /tɪ/ and /tʃ/, and /dɪ/ and /dʒ/ patterning together, with significantly higher

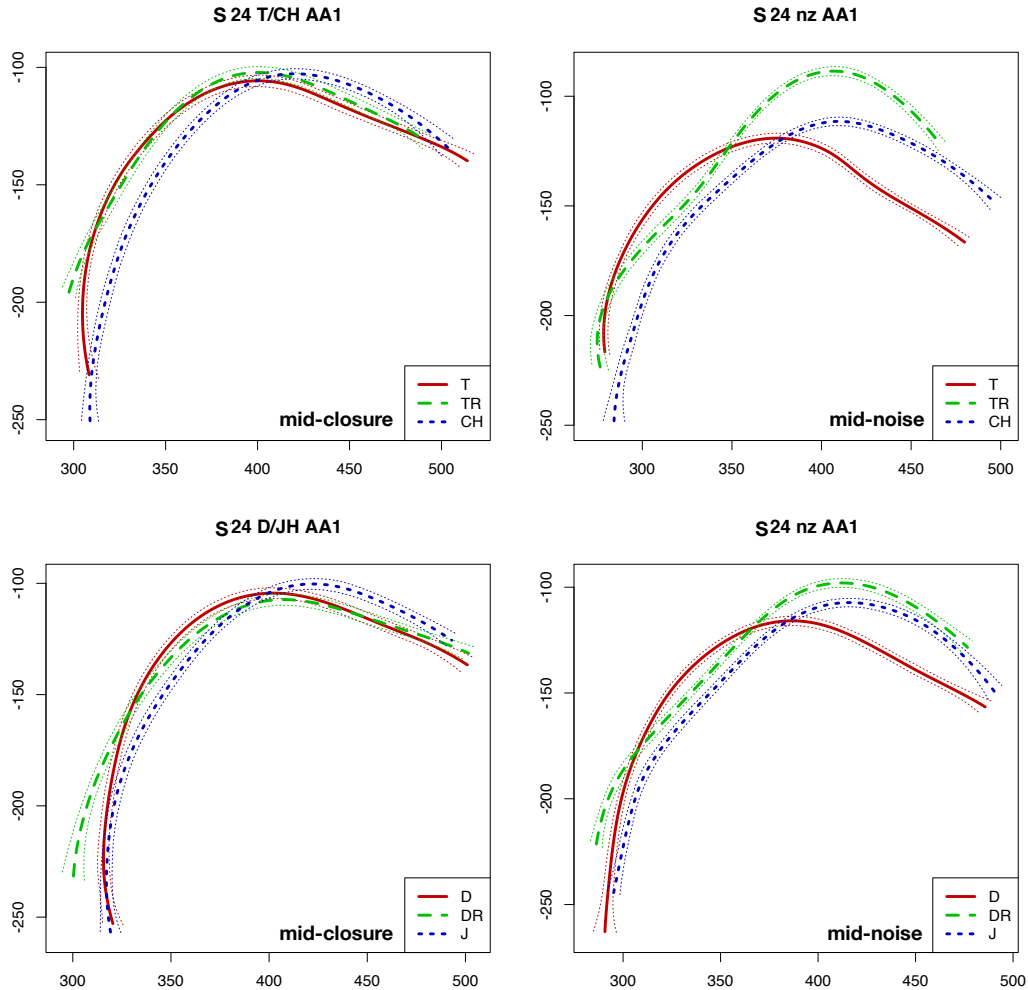
postures relative to /t/ and /d/ (TR,CH>T, DR,J>D).



**Figure 3.22:** S09: typical tongue contour patterns in /ʌ/ vowel contexts at the tongue root and tongue tip/blade in voiceless (upper, T TR CH) and voiced (lower, D DR J) contours, at the mid-point of the closure interval (left), and the mid-point of the noise interval (right).

In Figure 3.23, S24 exhibits typical tongue root patterns at both time points in /ɑ/ vowel contexts, but deviates from the group at the tongue tip/blade mid-closure (left), showing /tɪ/ and /t/, and /dɪ/ and /d/ patterning together, with relatively lower tongue tip/blade postures when compared with /tʃ/ and /dʒ/ (CH>TR,T, J>DR,D). At the mid-point of the noise interval (right), S24 shows three distinct tongue contours at the tongue tip/blade, with /tɪ/ and /dɪ/ exhibiting higher pos-

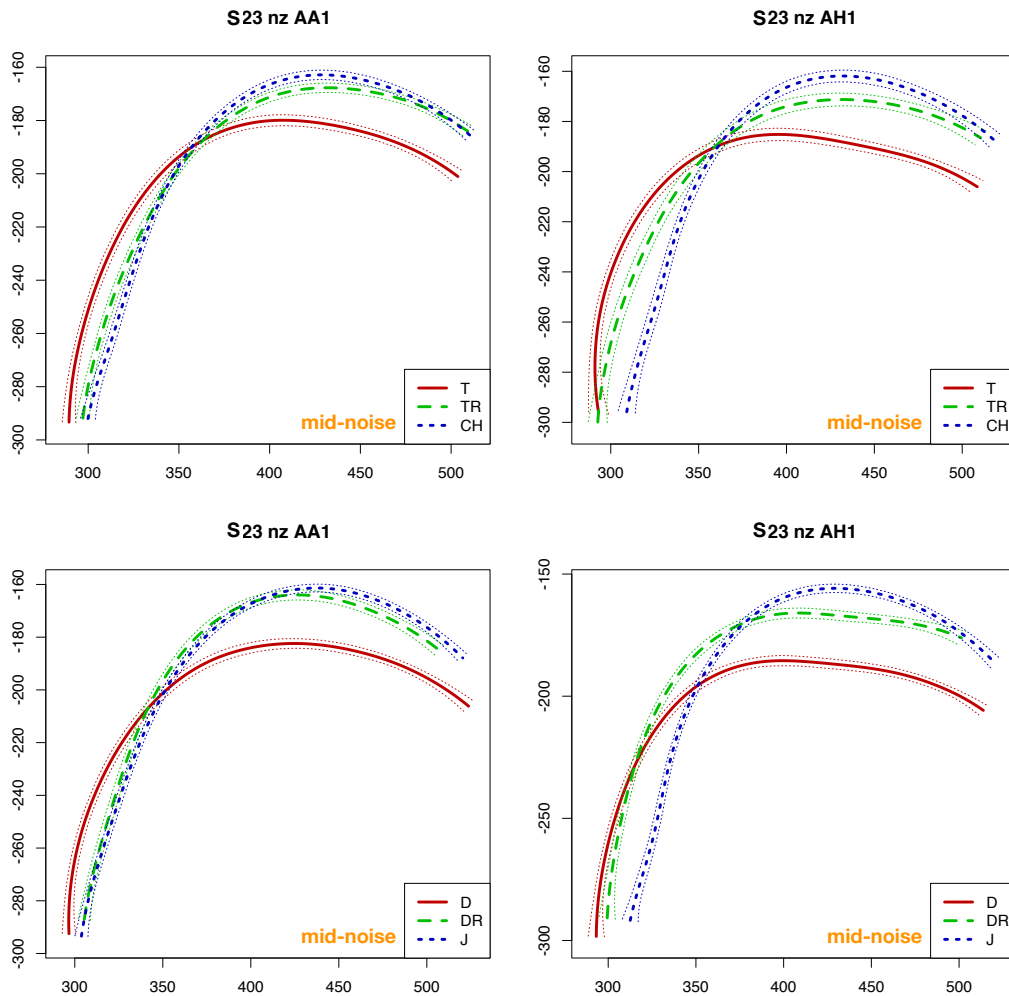
tures relative to /tʃ/ and /dʒ/, and /tʃ/ and /dʒ/ exhibiting higher postures relative to /t/ and /d/ (TR>CH>T, DR>J>D). At the tongue tip/blade in /ɑ/ vowel contexts (left), S23 also showed /tɪ/ and /tʃ/, and /dɪ/ and /dʒ/ patterning together, with significantly higher postures relative to /t/ and /d/ (TR,CH>T, DR,J>D).



**Figure 3.23:** S24: tongue contour patterns in /ɑ/ vowel contexts at the tongue root and tongue tip/blade in voiceless (upper, T TR CH) and voiced (lower, D DR J) contours, at the mid-point of the closure interval (left), and the mid-point of the noise interval (right).

Only one participant (S23) produced retroflex /ɪ/ postures, and only in /ʌ/ vowel contexts. At the mid-point of the closure interval, S23 exhibited typical patterns in both back vowel contexts, showing no significant differences at the tongue root

(TR=CH=T, DR=D=J), and showing /tɪ/ and /tʃ/, and /dɪ/ and /dʒ/ contours patterning together at the tongue tip/blade, with significantly higher postures relative to /t/ and /d/ (TR,T>CH, DR,D>J). In contrast, as shown in Figure 3.24, S23 deviated from the group at the mid-point of the noise interval in /ɑ/ vowel (left) and /ʌ/ vowel (right) contexts, exhibiting /t/ and /d/ postures that were relatively farther back than /tɪ/ and /tʃ/, and /dɪ/ and /dʒ/ (T>TR,CH, D>DR,J)<sup>9</sup>. In /ʌ/ vowel contexts (right), however, S23 exhibited /tɪ/ and /dɪ/ contours that were



**Figure 3.24:** S23: tongue contour patterns in /ɑ/ vowel (left) and /ʌ/ vowel (right) contexts at the mid-point of the noise interval (left and right), for voiceless (upper, T TR CH) and voiced (lower, D DR J) contours.

<sup>9</sup>Except voiced contours in the /ʌ/ vowel context, which exhibited a typical pattern: DR,D>J.

consistent with what we might expect to see with a more retroflex (tip-up) /ɪ/, and distinct from the other contours (CH>TR>T, J>DR>D).

### 3.2.2.4 Summary of observed patterns

The tables below provide a summary of participants' observed tongue root patterns (Table 3.3) and observed tongue tip/blade patterns (Table 3.4) at both points in time. Taken together, results show participants typically producing more retracted /tɪ/ and /dɪ/ tongue root postures in front vowel contexts, when compared with their relatively more advanced /t/ and /tʃ/, and /d/ and /dʒ/ tongue root postures (TR>CH,T, DR>D,J) at the mid-points of the closure and noise intervals. We expect more advanced tongue root postures in prevocalic /t tʃ d dʒ/, in anticipation of the tense high front vowel /i/ (produced with tongue root advancement), and tongue root retraction in pre-/ɪ/ contexts (produced with tongue root constriction at the pharyngeal wall). The observed results are consistent with these expected patterns. We do not anticipate tongue root advancement for the lax /ɪ/ vowel, so retracted /tɪ/ and /dɪ/ postures in this vowel context at the mid-point of the closure interval point to early production of the /ɪ/ gesture.

At the mid-point of the closure interval, participants exhibited no significant differences between contours at the tongue tip/blade in front vowel contexts, and while in /i/ vowel contexts this pattern persisted into the mid-point of the noise interval, in /ɪ/ vowel contexts, participants' /tɪ/ and /dɪ/ contours patterned with /tʃ/ and /dʒ/, showing higher tongue postures relative to /t/ and /d/. As the anterior portion of the tongue is less encumbered in this lax front vowel context relative to /i/, an articulatory pattern at the tongue tip/blade has more freedom to emerge.

| Context | Tongue Root Patterns | Closure Interval |        | Noise Interval |        |
|---------|----------------------|------------------|--------|----------------|--------|
|         |                      | Voiceless        | Voiced | Voiceless      | Voiced |
| /ɪ/     | 1. TR>T,CH; DR>D,J   | 9                | 9      | 12             | 12     |
|         | 2. TR=T=CH; DR=D=J   | 3                | 3      | –              | –      |
|         | 3. TR,T>CH; DR,D>J   | –                | –      | –              | –      |
|         | 4. T>TR,CH; D,DR,J   | –                | –      | –              | –      |
| /i/     | 1. TR>T,CH; DR>D,J   | 10               | 10     | 11             | 11     |
|         | 2. TR=T=CH; DR=D=J   | 1                | 2      | 1              | 1      |
|         | 3. TR,T>CH; DR,D>J   | 1                | –      | –              | –      |
|         | 4. T>TR,CH; D,DR,J   | –                | –      | –              | –      |
| /ʌ/     | 1. TR>T,CH; DR>D,J   | 2                | 5      | 4              | 4      |
|         | 2. TR=T=CH; DR=D=J   | 5                | 5      | 1              | 4      |
|         | 3. TR,T>CH; DR,D>J   | 5                | 2      | 6              | 3      |
|         | 4. T>TR,CH; D,DR,J   | –                | –      | 1              | 1      |
| /ɑ/     | 1. TR>T,CH; DR>D,J   | 2                | 3      | 1              | 2      |
|         | 2. TR=T=CH; DR=D=J   | 4                | 6      | 4              | 4      |
|         | 3. TR,T>CH; DR,D>J   | 5                | 3      | 5              | 4      |
|         | 4. T>TR,CH; D,DR,J   | 1                | –      | 2              | 2      |

**Table 3.3:** Summary of observed **tongue root patterns** by vowel context for voiceless and voiced contours from the mid-points of the closure and noise intervals, where > represents greater distance from the point of origin, reflected in each SSANOVA as a tongue root contour with a smaller value along the  $x$ -axis.

In back vowel contexts, participants showed more variation at the tongue root, frequently showing no significant differences between contours. While several participants exhibited retracted /tɪ/ and /dɪ/ postures relative to the other two contours at both time points in both back vowel contexts, participants' /tɪ/ and /t/, and /dɪ/ and /d/ contours frequently patterned together. We expect more retracted tongue root postures in pre-/ɪ/ contexts (due to anticipatory constriction at the pharyngeal wall), but tongue root retraction is also expected in back vowel contexts (particularly /ɑ/). It seems plausible that /tɪ/ and /t/, and /dɪ/ and /d/ contours were patterning together early in back vowel contexts (relative to /tʃ/ and /dʒ/) for different reasons: /ɪ/ is an articulatorily complex sound to produce and there are no articulatory consequences of producing it early, and /t/ and /d/ are shorter in

duration than affricates (Byrd, 1993), so can begin co-production of the following back vowel (produced with a retracted tongue root) earlier in time relative to /tʃ/ and /dʒ/. At the tongue tip/blade, participants typically showed /tɹ/ and /tʃ/, and /dɹ/ and /dʒ/ contours patterning together. These patterns were more visible in back vowel contexts, where the anterior portion of the tongue is articulatorily less encumbered than in front vowel contexts.

| Context | Tongue Tip/Blade Patterns | <i>Closure Interval</i> |        | <i>Noise Interval</i> |        |
|---------|---------------------------|-------------------------|--------|-----------------------|--------|
|         |                           | Voiceless               | Voiced | Voiceless             | Voiced |
| /ɪ/     | 1. TR,CH>T & DR,J>D       | 3                       | 2      | 10                    | 8      |
|         | 2. TR=T=CH & DR=D=J       | 9                       | 10     | 2                     | 4      |
|         | 3. CH,T>TR & DR,D>DR      | –                       | –      | –                     | –      |
|         | 4. CH>TR,T & DR>DR,D      | –                       | –      | –                     | –      |
|         | 5. CH>TR>T & J>DR>D       | –                       | –      | –                     | –      |
|         | 6. TR>CH>T & DR>J>D       | –                       | –      | –                     | –      |
| /i/     | 1. TR,CH>T & DR,J>D       | –                       | –      | –                     | –      |
|         | 2. TR=T=CH & DR=D=J       | 10                      | 11     | 9                     | 10     |
|         | 3. CH,T>TR & DR,D>DR      | 2                       | 1      | 3                     | 2      |
|         | 4. CH>TR,T & DR>DR,D      | –                       | –      | –                     | –      |
|         | 5. CH>TR>T & J>DR>D       | –                       | –      | –                     | –      |
|         | 6. TR>CH>T & DR>J>D       | –                       | –      | –                     | –      |
| /ʌ/     | 1. TR,CH>T & DR,J>D       | 6                       | 3      | 7                     | 7      |
|         | 2. TR=T=CH & DR=D=J       | 5                       | 7      | –                     | –      |
|         | 3. CH,T>TR & DR,D>DR      | –                       | –      | –                     | –      |
|         | 4. CH>TR,T & J>DR,D       | 1                       | 1      | –                     | –      |
|         | 5. CH>TR>T & J>DR>D       | –                       | 1      | 1                     | 1      |
|         | 6. TR>CH>T & DR>J>D       | –                       | –      | 4                     | 4      |
| /ɑ/     | 1. TR,CH>T & DR,J>D       | 9                       | 8      | 8                     | 9      |
|         | 2. TR=T=CH & DR=D=J       | 2                       | 2      | –                     | 1      |
|         | 3. CH,T>TR & DR,D>DR      | –                       | –      | –                     | –      |
|         | 4. CH>TR,T & DR>DR,D      | 1                       | 1      | –                     | –      |
|         | 5. CH>TR>T & J>DR>D       | –                       | 1      | –                     | –      |
|         | 6. TR>CH>T & DR>J>D       | –                       | –      | 4                     | 2      |

**Table 3.4:** Summary of observed **tongue tip/blade patterns** by vowel context for voiceless and voiced contours from the mid-points of the closure and noise intervals, where > represents greater distance from the point of origin, reflected in each SSANOVA as a tongue tip/blade contour with a smaller negative value along the *y*-axis.

### 3.2.2.5 Comparing across tongue curves

In the figures that follow (Figures 3.25 to 3.28), distances between the SSANOVA fit curves for voiceless (T, TR, CH) contours presented above are compared for each participant across various contexts and plotted relative to one another<sup>10</sup>. Distances within specific regions along the curve are calculated using angles determined from polar coordinate data (radians) for each curve. For example, each sub-figure in Figure 3.25 presents between-participant comparisons of tongue root differences at the mid-point of the closure interval, in one of four vowel contexts: /ɪ i ʌ α/ (Figure 3.26 examines this same data at the mid-point of the noise interval following release of the closure). The *y*-axis for each sub-figure shows the difference between each participant's /tɹ/ and /tʃ/ tongue roots, determined based on the maximum difference between these two curves in the tongue root interval (defined using polar coordinates as the region between 0 and  $\pi/4$ ). The *x*-axis shows the difference between each participant's /tɹ/ and /t/ tongue roots. Tongue root retraction of /tɹ/ (relative to the sounds shown on each axis) is represented as a positive value, while a negative value indicates the tongue root of one of the other two contours is more retracted, and a value of 0 indicates no difference between the two contours. In contrast, each sub-figure in Figure 3.27 presents between-participant comparisons of differences at the tongue tip/blade at the mid-point of the closure interval, in one of four vowel contexts: /ɪ i ʌ α/ (Figure 3.28 examines this same data at the mid-point of the noise interval following release of the closure). In these plots, the *y*-axis for each sub-figure shows the maximum difference between each participant's /tɹ/ and /tʃ/ at the tongue tip/blade (defined as the region between  $\pi/2$  and  $\pi/4$ ), while the *x*-axis shows the maximum difference between each participant's /tɹ/ and /t/ at the tongue tip/blade. A positive value along each axis represents a higher

---

<sup>10</sup>As illustrated in the tongue contour (SSANOVA) results reported in the preceding sections, participants exhibited similar tongue postures in /tɹ/ and /dɹ/ sequences. While the scatterplots that follow present comparisons of voiceless (CH, TR, T) tokens only, these overall patterns were observed in comparisons of voiced (J, DR, D) tokens as well.

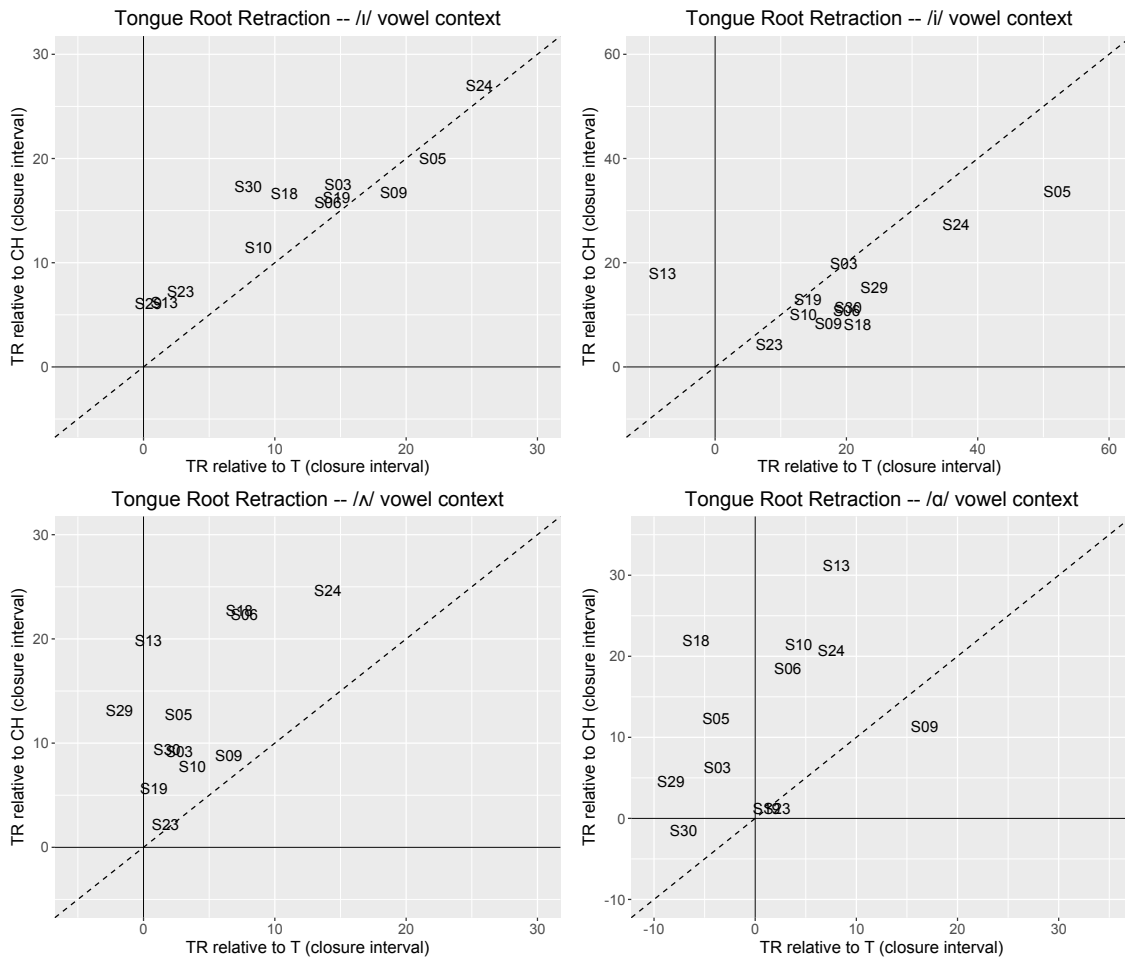
tongue tip/blade position relative to the other contour being compared.

Compatible with both coproduction hypotheses in Figure 3.1 (a, b), participants consistently exhibited early tongue root retraction in /tɹ/ and /dɹ/ sequences at the mid-points of the closure interval (Figure 3.25) and the noise interval following release (Figure 3.26)<sup>11</sup>. The gestural scores made no predictions about following vowels, but vowel context was relevant in terms of how prevocalic consonants patterned with /tɹ/ and /dɹ/ sequences. Participants exhibited the greatest /tɹ/ and /dɹ/ tongue root retraction relative to the other two contours in front vowel contexts at both points in time. While there was no evidence of /tɹ/ and prevocalic /tʃ/, or /dɹ/ and prevocalic /dʒ/ exhibiting similar tongue root retraction in any vowel context at either time point, /tɹ/ and prevocalic /t/, and /dɹ/ and prevocalic /d/ frequently patterned together in back vowel contexts (Figures 3.25 and 3.26, bottom, *x*-axis values close to 0). Several participants showed more retracted prevocalic /t/ and /d/ postures in back vowel<sup>12</sup> contexts relative to /tɹ/ and /dɹ/ at the mid-point of the closure interval (S03, S05, S18, S29, S30) and the mid-point of the noise interval (S05, S23, S30).

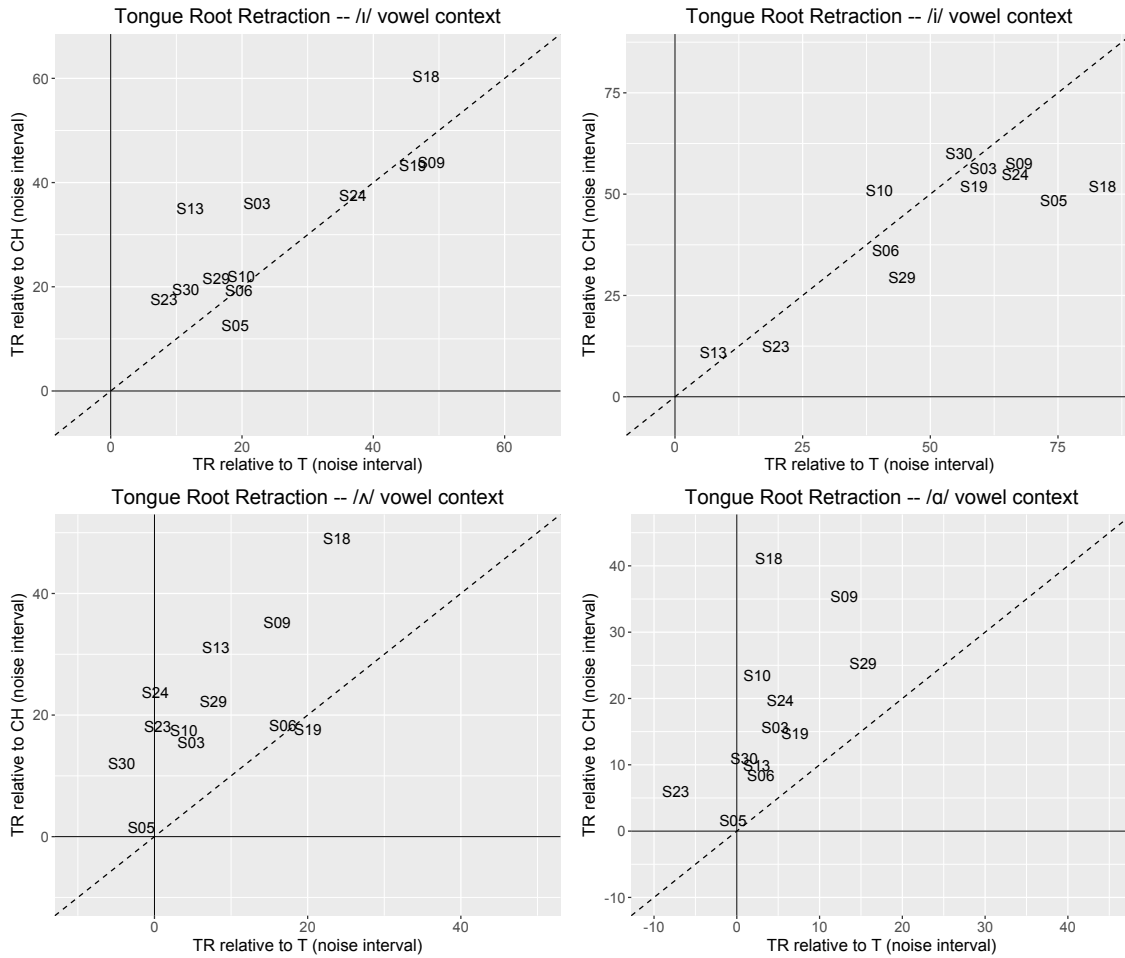
---

<sup>11</sup>A note about Figures 3.25 through 3.28: axis scales are different across vowel contexts, in order to adequately capture the distribution. Values are higher where differences between contours are greater, resulting in broader axis ranges, as reflected in the plots.

<sup>12</sup>For one participant (S13), this pattern was observed in /i/ vowel contexts of voiceless tokens at the mid-point of the closure interval as well (Figure 3.25, top right).



**Figure 3.25:** At the mid-point of the closure interval, tongue root retraction of /tɪ/ (TR), across front vowel (top left: /ɪ/, top right: /i/) and back vowel (bottom left: /ʌ/, bottom right: /ɑ/) contexts, relative to prevocalic /t/ (T) and /tʃ/ (CH). Positive values ( $> 0$ ) indicate /tɪ/ tongue root retraction relative to /tʃ/ (y-axis), and /t/ (x-axis).



**Figure 3.26:** At the mid-point of the noise interval, tongue root retraction of /tɹ/ (TR) across front vowel (top left: /ɪ/, top right: /i/) and back vowel (bottom left: /ʌ/, bottom right: /ɑ/) contexts, relative to prevocalic /t/ (T) and /tʃ/ (CH). Positive values ( $> 0$ ) indicate /tɹ/ tongue root retraction relative to /tʃ/ ( $y$ -axis), and /t/ ( $x$ -axis).

Consistent with hypothesis (b), at the tongue tip/blade at the mid-point of the closure interval, participants' articulations of /tɹ/ and /dɹ/ typically patterned with /tʃ/ and /dʒ/ in back vowel contexts (Figure 3.27, bottom, values close to 0 on the  $y$ -axis), showing more [tʃ]-like and [dʒ]-like postures that were significantly higher than those observed with prevocalic /t/ and /d/. For some participants (e.g. S30 in the /ʌ/ vowel context), /tɹ/ and /dɹ/ contours were lower at the tongue tip/blade than /tʃ/ and /dʒ/ (Figure 3.27, bottom, negative values ( $< 0$ ) on the  $y$ -axis). In front vowel contexts, there were frequently no statistically significant differences

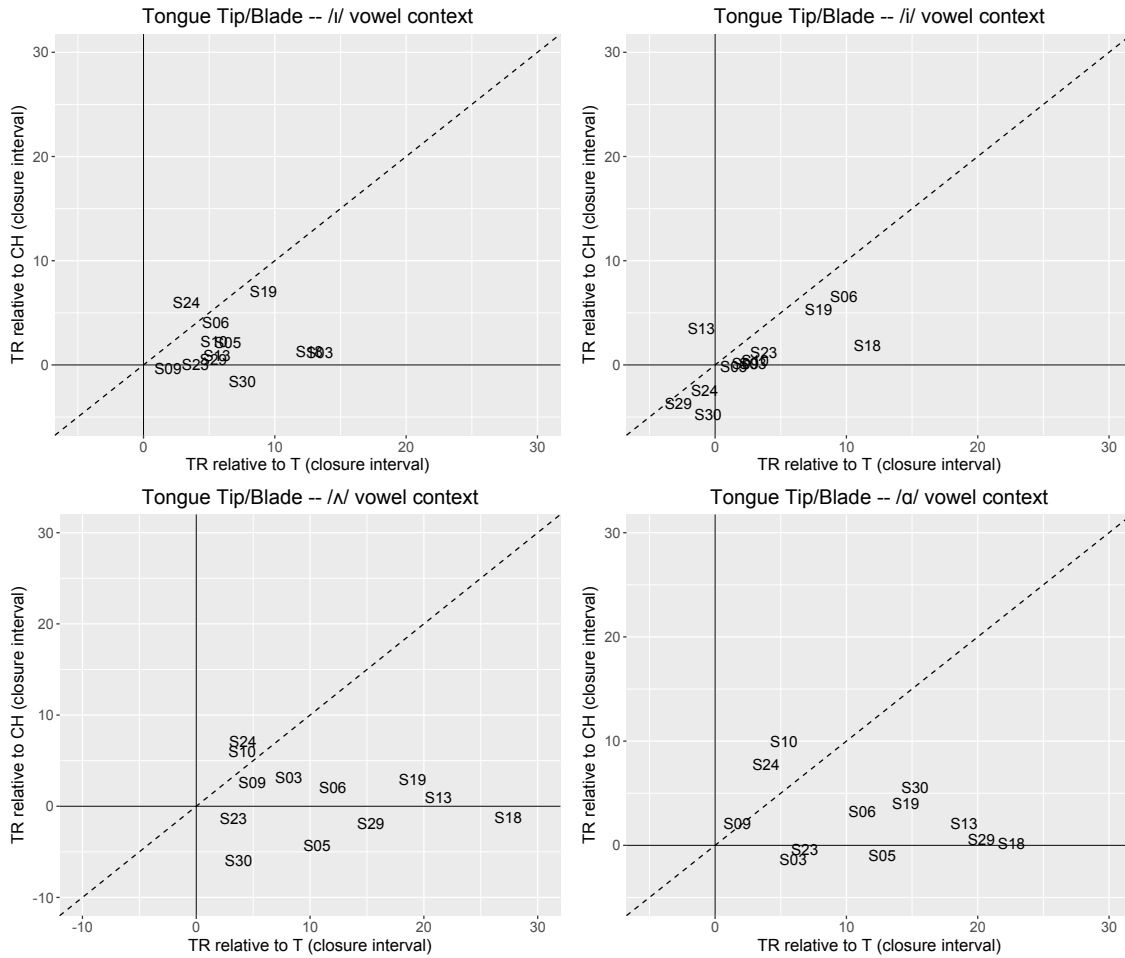
between contours at the tongue tip/blade (Figure 3.27, top, values close to 0,0 along the  $x$  and  $y$  axes), though some participants exhibited lower /tɹ/ and /dɹ/ tongue tip/blade postures relative to the other two contours (S05, S13, S24, S29, S30), or showed /tɹ/ and /dɹ/ tip/blade postures that patterned with /tʃ/ and /dʒ/ (S03, S05, S18, S19), as observed in back vowel contexts.

By the mid-point of the noise interval following release, participants consistently exhibited [tʃ]-like and [dʒ]-like postures in /tɹ/ and /dɹ/ sequences that were significantly higher than prevocalic /t/ and /d/ contours at the tongue tip/blade across /ɪ ʌ ɑ/ vowel contexts (Figure 3.28). Tongue contours in /i/ vowel contexts were the exception, with participants showing no significant differences between contours, or exhibiting /tɹ/ and /dɹ/ contours that were significantly lower than prevocalic /t/ and /d/.

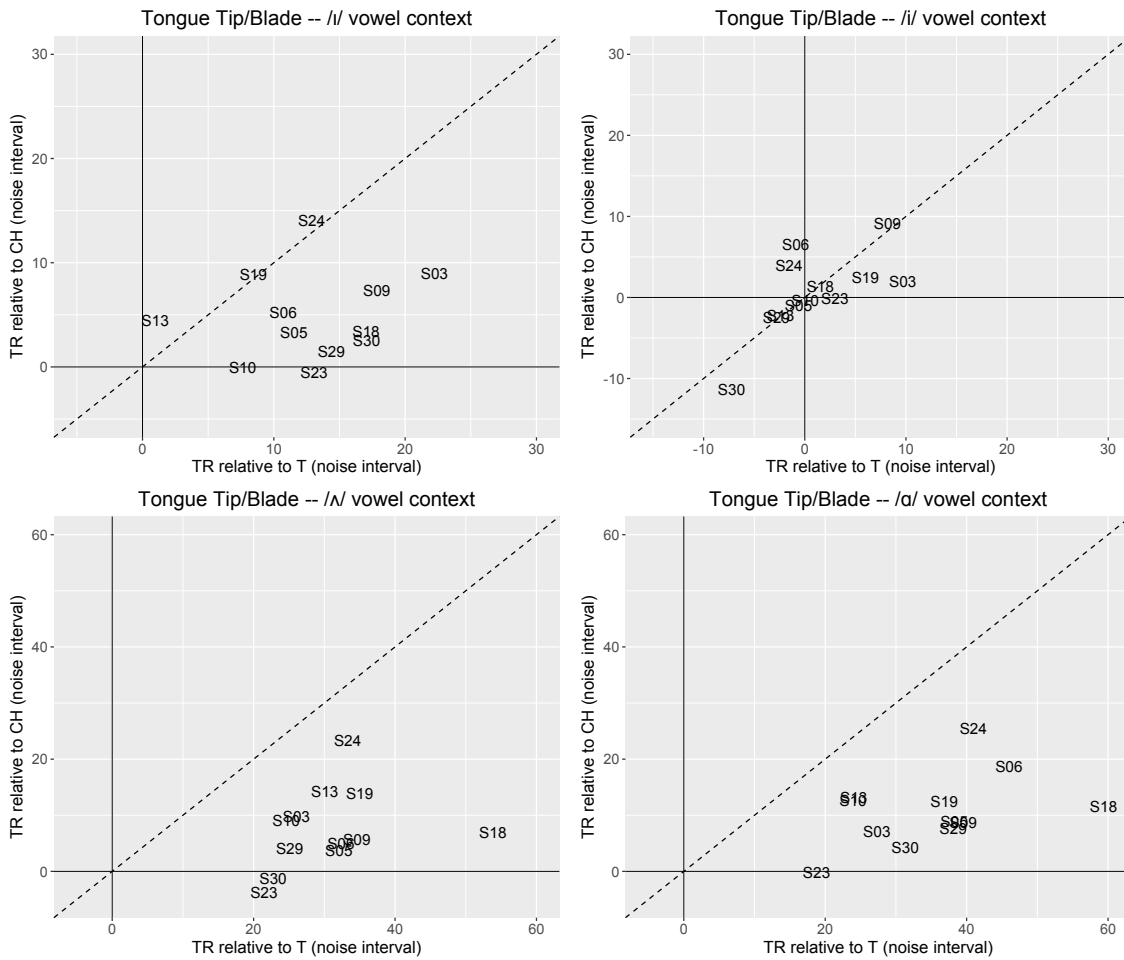
In both hypotheses (a) and (b), [ɹ]-like tongue postures are predicted by the mid-point of the noise interval, as the tongue is articulatorily less encumbered following release, and free to produce the gesture for the /ɹ/. Given that [ʃ] and bunched [ɹ] both involve tongue blade raising with post-alveolar constriction, it was reasoned that participants' articulations of /tɹ/ and /dɹ/ may look articulatorily similar to [tʃ] and [dʒ] at the tongue tip/blade under either scenario. As described above, one possible way to disambiguate a [tʃ] or [dʒ] from an [ɹ] is to look for concavity at the tongue dorsum (an articulatory characteristic of English bunched [ɹ]). As illustrated in Figure 3.29 (S18), while participants typically exhibited /tɹ/ and /tʃ/ and /dɹ/ and /dʒ/ tongue tip/blade contours that patterned together at the mid-point of the noise interval, almost all participants (S03, S05, S06, S09, S13, S18, S19, S24, S29, S30) showed some concavity at the tongue dorsum in /tɹ/ and /dɹ/ sequences as well, suggesting early bunched [ɹ] postures and consistent with the hypotheses<sup>13</sup>.

---

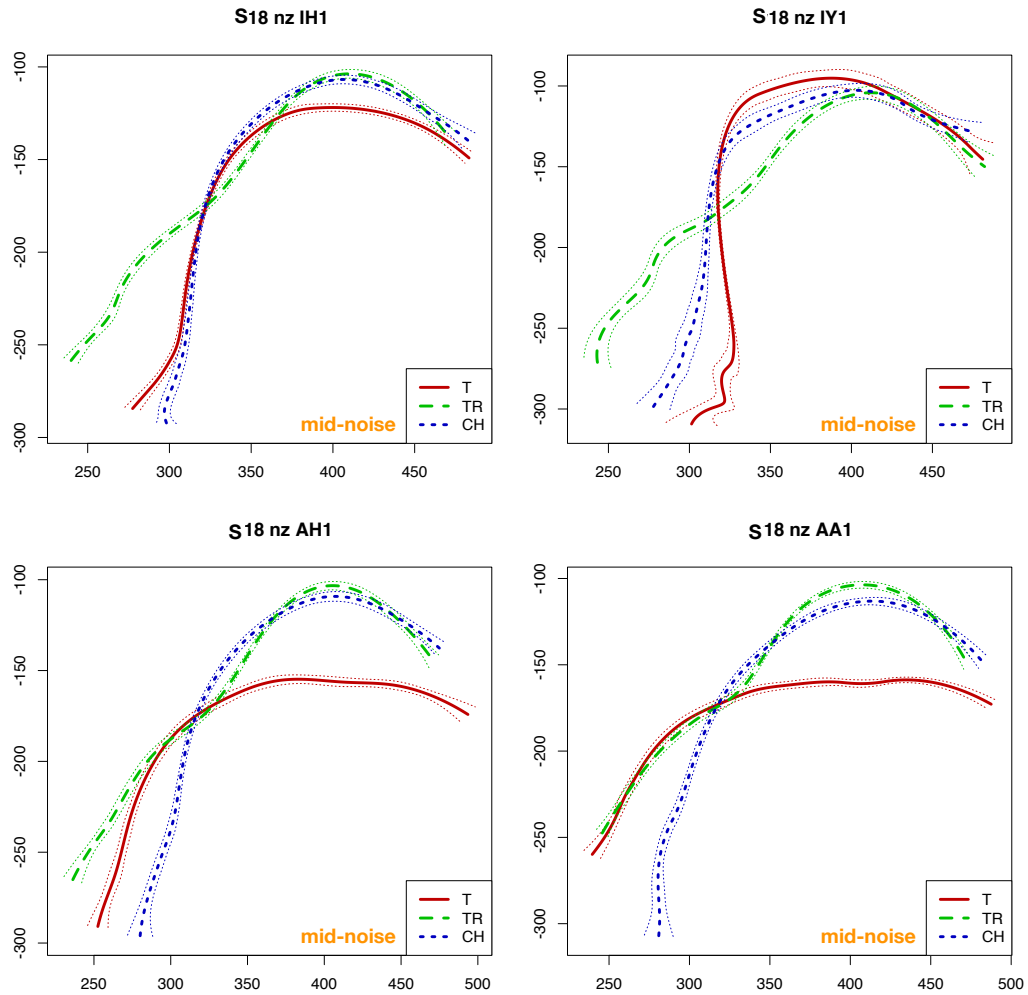
<sup>13</sup>S23's /tɹ/ and /dɹ/ postures in back vowel contexts showed no concavity at the tongue dorsum, but did exhibit retroflex /ɹ/ postures, which is also consistent with the hypotheses.



**Figure 3.27:** At the mid-point of the closure interval, tongue tip/blade height of /tɪ/ (TR) across front vowel (top left: /ɪ/, top right: /i/) and back vowel (bottom left: /ʌ/, bottom right: /ɑ/) contexts, relative to pre-vocalic /t/ (T) and /tʃ/ (CH). Positive values (> 0) indicate higher /tɪ/ tongue tip/blade relative to /tʃ/ (y-axis), and /t/ (x-axis).



**Figure 3.28:** At the mid-point of the noise interval, tongue tip/blade height of /tɹ/ (TR) across front vowel (top left: /ɪ/, top right: /i/) and back vowel (bottom left: /ʌ/, bottom right: /ɑ/) contexts, relative to prevocalic /t/ (T) and /tʃ/ (CH). Positive values (> 0) indicate higher /tɹ/ tongue tip/blade relative to /tʃ/ (*y*-axis), and /t/ (*x*-axis).

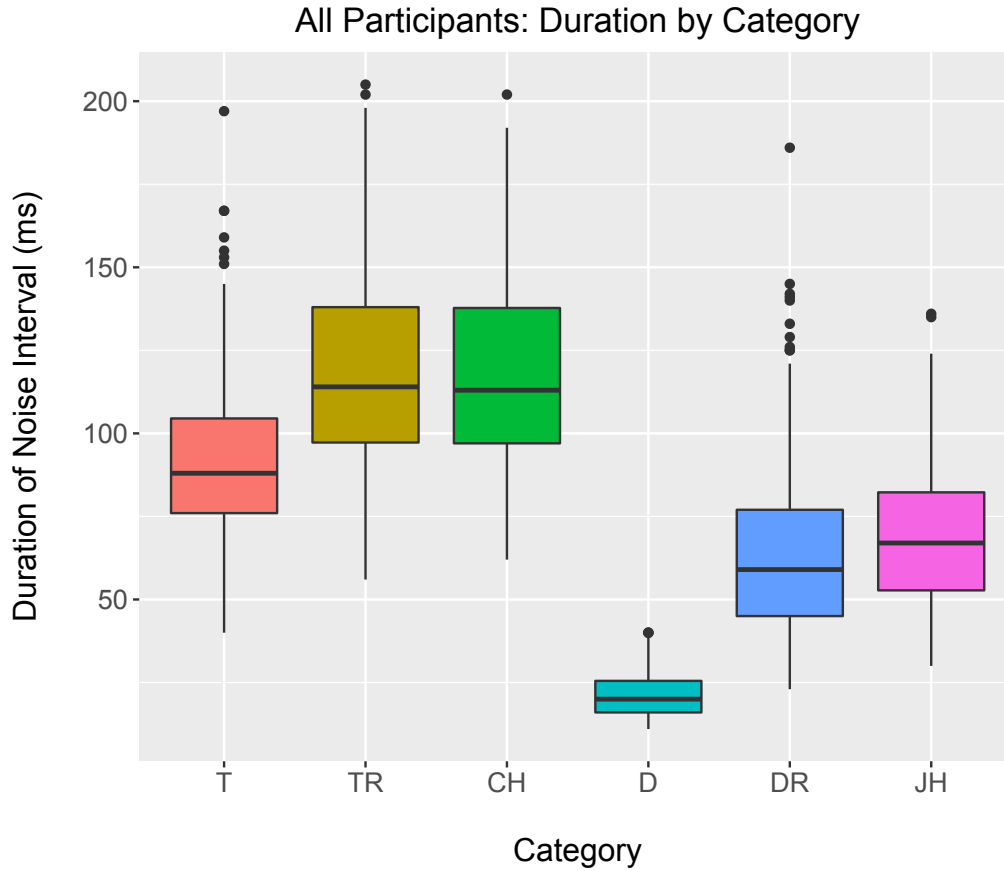


**Figure 3.29:** S18 showing concavity at the tongue dorsum at the mid-point of the noise interval in /tɪ/ contours across vowel contexts: /ɪ/ (upper left), /i/ (upper right), /ʌ/ (lower left), /ɑ/ (lower right).

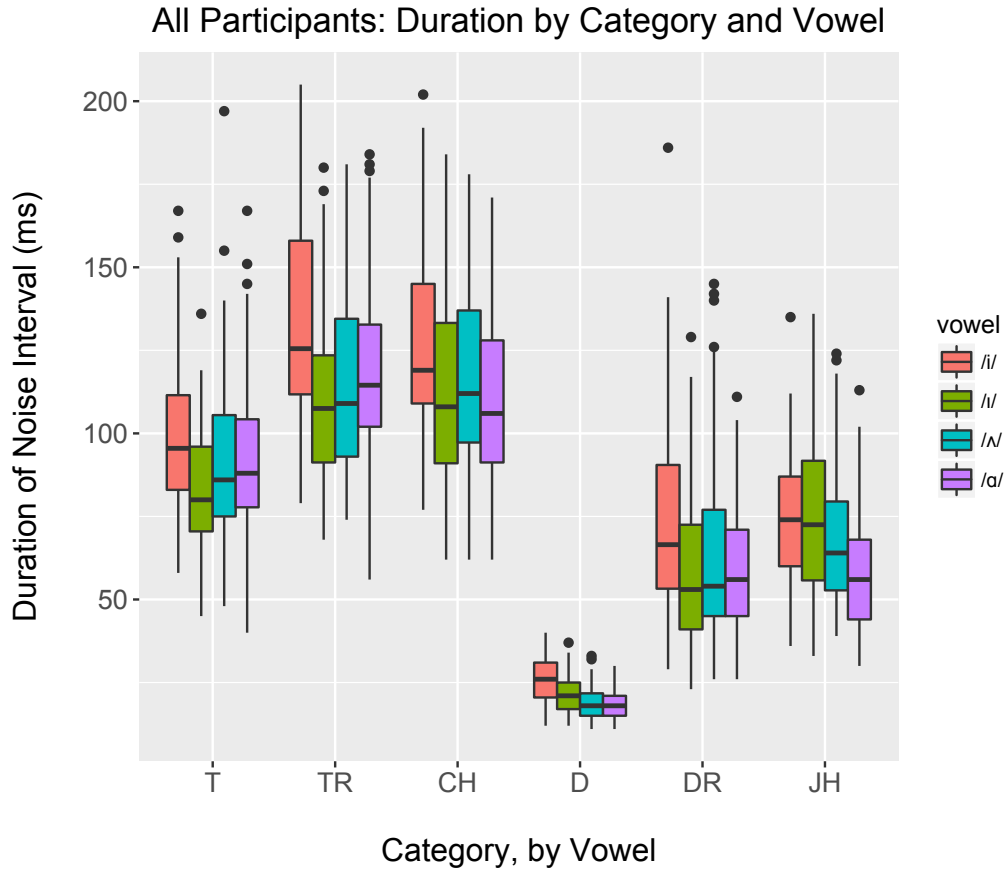
### 3.2.3 Duration

If participants in the Production Study are producing affricated variants of /t/ and /d/ before /ɹ/, it is reasonable to expect that these variants may be similar in duration to phonological affricates, /tʃ/ and /dʒ/. In addition, Ohala and Solé (2010) argue that stops show an extended period of turbulence following release into a segment with a high tongue position. The tense high front vowel /i/, as well as /ɹ/, are produced with high tongue postures and may result in what Ohala and Solé (2010:47) refer to as “emergent affricates”, with an extended period of turbulence that may be perceived as affrication. In other words, if the ‘affrication’ in /tɹ/ and /dɹ/ sequences is due to aerodynamic factors, it seems plausible that /t/ and /d/ before [i] may exhibit similar duration to /t/ and /d/ before [ɹ].

In the figures that follow, duration of the noise interval following the release of closure corresponds to VOT for /t/ and /d/, and (roughly) to the interval of frication for /tʃ/ and /dʒ/. Figure 3.30 (all participants) presents duration (in milliseconds) of the noise interval by category for /t d tʃ dʒ/ before vowels (T, D, CH, JH), and /t d/ before /ɹ/ (TR, DR). Voiceless stops before vowels exhibited longer duration following closure than voiced stops before vowels (T>D). Affricates before vowels showed greater duration than stops before vowels (CH>T, JH>D). Stops before /ɹ/ showed greater duration than stops before vowels (TR>T, DR>D), and were similar in duration to affricates (TR≈CH, DR≈JH). Broken down by vowel context (Figure 3.31, all participants), noise intervals across categories were longest before /i/, consistent with Ohala and Solé (2010), but the shortest noise interval for /tɹ/ and /dɹ/ was still longer than the longest pre-[i] noise interval for /t/ and /d/.



**Figure 3.30:** All participants: duration (ms) of noise interval by category: /t/ before vowels (T), /t/ before /ɪ/ (TR), /tʃ/ before vowels (CH), /d/ before vowels (D), /d/ before /ɪ/ (DR), and /dʒ/ before vowels (JH). (Boxes represent interquartile range, solid horizontal lines in boxes represent median, whiskers represent range of data points falling within 1.5 times the interquartile range; circles depict outliers.)



**Figure 3.31:** All participants: duration (ms) of noise interval by category (T, TR, CH, D, DR, JH) and vowel context (/i/ /ɪ/, /ʌ/, and /ɑ/). (Boxes represent interquartile range, solid horizontal lines in boxes represent median, whiskers represent range of data points falling within 1.5 times the interquartile range; circles depict outliers.)

### 3.2.3.1 Statistical modeling

The *lmer* function in the lme4 package (R Development Core Team, 2008) was used to generate mixed-effects linear regression models with Duration as the dependent variable. The models presented below<sup>14</sup> include the same random intercepts for SPEAKER and WORD. The significance threshold is defined as a *t* value of greater than 2.0 or less than -2.0.

Model 1a and Model 1b look at Duration by CATEGORY, (duration ~ category +

<sup>14</sup>I considered using a random slopes structure for each word by speaker (1 + word|speaker), but these models resulted in a significant increase in AIC over reported models.

(1 | speaker) + (1 | word)), in both voiceless (1a) and voiced (1b) conditions. Model 1a (Table 3.5) indicates a significant effect of Category /t/ vs. the reference level of /tɹ/ ( $t$  value: -6.595), but no significant effect of Category /tʃ/ vs. the reference level of /tɹ/ ( $t$  value: -0.378) in the voiceless condition, with /tɹ/ exhibiting significantly longer duration of the noise interval when compared with /t/, but no significant difference in duration when compared with /tʃ/. Similarly, Model 1b (Table 3.6) shows a significant effect of Category /d/ vs. the reference level of /dɹ/ ( $t$  value: -14.293), but no significant effect of Category /dʒ/ vs. the reference level of /dɹ/ ( $t$  value: 1.517), with /dɹ/ exhibiting significantly longer duration of the noise interval when compared with /d/, but no significant difference in duration when compared with /dʒ/.

| AIC: 6732.9   | Estimate | Std. Error | $t$ value |
|---------------|----------|------------|-----------|
| (Intercept)   | 119.029  | 6.483      | 18.361    |
| Category /t/  | -27.587  | 4.183      | -6.595    |
| Category /tʃ/ | -1.582   | 4.180      | -0.378    |

**Table 3.5:** Summary of Duration Model 1a (TR):  
 $\text{duration} \sim \text{category} + (1 | \text{subject}) + (1 | \text{word})$

| AIC: 6457.4   | Estimate | Std. Error | $t$ value |
|---------------|----------|------------|-----------|
| (Intercept)   | 64.440   | 4.457      | 14.458    |
| Category /d/  | -43.643  | 3.053      | -14.293   |
| Category /dʒ/ | 4.632    | 3.053      | 1.517     |

**Table 3.6:** Summary of Duration Model 1b (DR):  
 $\text{duration} \sim \text{category} + (1 | \text{subject}) + (1 | \text{word})$

Model 2a and Model 2b look at Duration by CATEGORY and VOWEL, ( $\text{duration} \sim \text{category} + \text{vowel} + (1 | \text{speaker}) + (1 | \text{word})$ ), in both voiceless (2a) and voiced (2b) conditions. Model 2a (Table 3.7) shows the same significant effect of Category as in Model 1a, and also a significant effect of Vowel /ɪ/ vs. the reference level of /i/

( $t$  value: -5.591), Vowel /ʌ/ vs. the reference level of /i/ ( $t$  value: -4.321), and Vowel /ɑ/ vs. the reference level of /i/ ( $t$  value: -4.375), with voiceless tokens exhibiting significantly longer noise interval duration across categories before /i/ than before the other three (/ɪ ʌ ɑ/) vowels.

Similarly, Model 2b (Table 3.8) shows the same significant effect of Category as Model 1b, and also a significant effect of Vowel /ɪ/ vs. the reference level of /i/ ( $t$  value: -2.283), Vowel /ʌ/ vs. the reference level of /i/ ( $t$  value: -2.836), and Vowel /ɑ/ vs. the reference level of /i/ ( $t$  value: -5.114), with voiced tokens exhibiting significantly longer noise interval duration across categories before /i/ than before the other three (/ɪ ʌ ɑ/) vowels.

| AIC: 6717.1   | Estimate | Std. Error | $t$ value |
|---------------|----------|------------|-----------|
| (Intercept)   | 129.782  | 6.319      | 20.540    |
| Category /t/  | -27.618  | 2.614      | -10.565   |
| Category /tʃ/ | -1.539   | 2.609      | -0.590    |
| Vowel /ɪ/     | -16.812  | 3.007      | -5.591    |
| Vowel /ʌ/     | -13.022  | 3.014      | -4.321    |
| Vowel /ɑ/     | -13.211  | 3.020      | -4.375    |

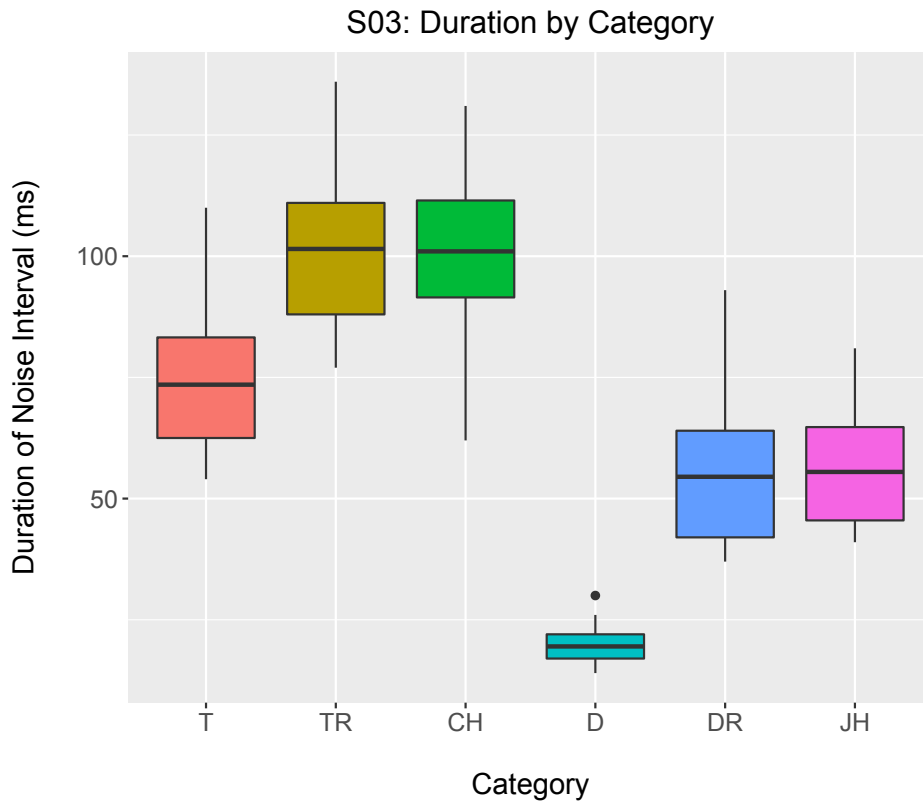
**Table 3.7:** Summary of Duration Model 2a (TR):  
 $\text{duration} \sim \text{category} + \text{vowel} + (1 \mid \text{subject}) + (1 \mid \text{word})$

| AIC: 6445.8   | Estimate | Std. Error | $t$ value |
|---------------|----------|------------|-----------|
| (Intercept)   | 70.580   | 4.405      | 16.022    |
| Category /d/  | -43.649  | 2.0810     | -20.984   |
| Category /dʒ/ | 4.643    | 2.080      | 2.232     |
| Vowel /ɪ/     | -5.479   | 2.400      | -2.283    |
| Vowel /ʌ/     | -6.809   | 2.401      | -2.836    |
| Vowel /ɑ/     | -12.338  | 2.413      | -5.114    |

**Table 3.8:** Summary of Duration Model 2b (DR):  
 $\text{duration} \sim \text{category} + \text{vowel} + (1 \mid \text{subject}) + (1 \mid \text{word})$

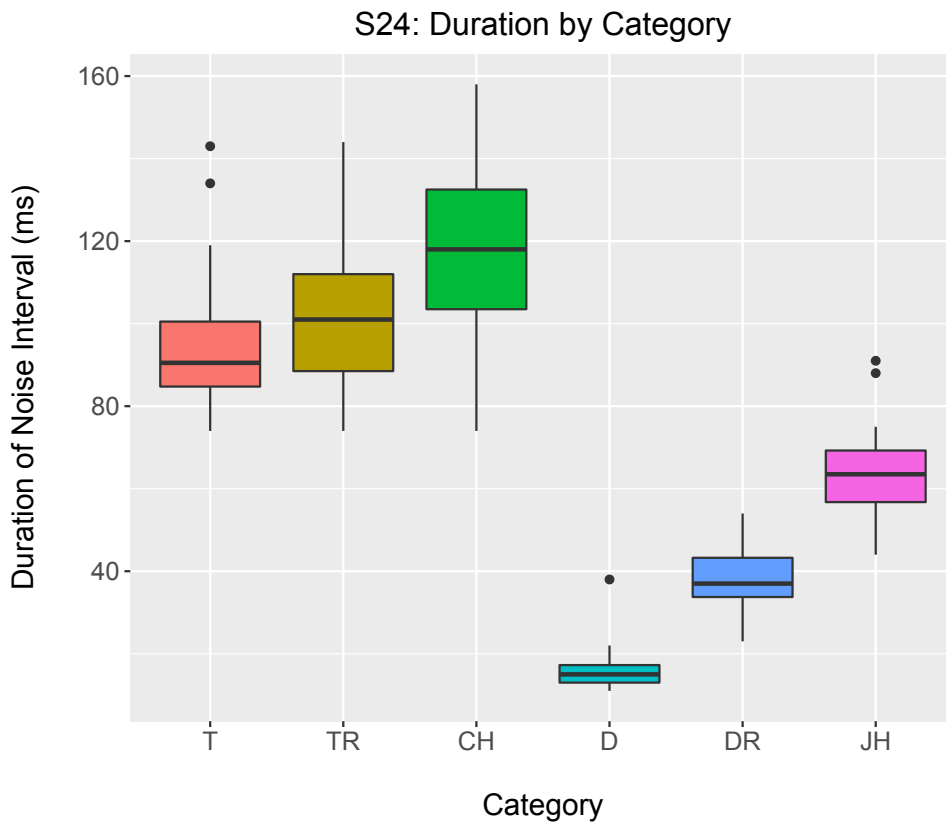
Although a small significant difference was observed at the group level between DR and JH, the overall pattern ( $D < TR, DR \approx CH, JH$ ) held, with /dɪ/ and /dʒ/ pat-

turning together, relative to /d/. Overall noise interval duration patterns remained stable at the level of the participant. As illustrated in Figure 3.32, S03's affricates before vowels showed greater duration than stops before vowels (CH, JH > T, D). S03's stops before /ɹ/ showed greater duration than stops before vowels (TR, DR > T, D), and were similar in duration to affricates (TR, DR ≈ CH, JH) before vowels. In a one-way ANOVA, the main effect of category was significant in S03's voiceless [ $F_{2,62} = 20.76, p < .001$ ] and voiced [ $F_{2,67} = 76.74, p < .001$ ] contexts. Post hoc comparisons using the Tukey HSD test showed significant duration differences between T and TR ( $p < .001$ ), and T and CH ( $p < .001$ ), but no significant difference between TR and CH ( $p = 0.99$ ), and significant differences between D and DR ( $p < .001$ ), and D and JH ( $p < .001$ ), but no significant difference between DR and JH ( $p = .97$ ).



**Figure 3.32:** S03 exhibiting the typical (group) noise interval duration pattern (T, D < TR, DR ≈ CH, JH). (Boxes represent interquartile range, solid horizontal lines in boxes represent median, whiskers represent range of data points falling within 1.5 times the interquartile range; circles depict outliers.)

S24 was the exception to the group level pattern, showing duration differences across categories in voiceless [ $F_{2,67} = 9.267, p < .001$ ] and voiced [ $F_{2,69} = 176.7, p < .001$ ] contexts (Figure 3.33), with affricates before vowels exhibiting significantly longer duration than stops before /ɹ/ (CH,JH>TR,DR). This atypical duration pattern is consistent with S24’s articulatory results at the mid-point of the noise interval, which show three distinctly different tongue curves before /ɹ/ relative to tongue curves for prevocalic /t/ and /d/, and /tʃ/ and /dʒ/ (see, e.g., Figure 3.23).



**Figure 3.33:** S24 deviating from the typical (group) noise interval duration pattern (T,D<TR,DR<CH,JH). (Boxes represent interquartile range, solid horizontal lines in boxes represent median, whiskers represent range of data points falling within 1.5 times the interquartile range; circles depict outliers.)

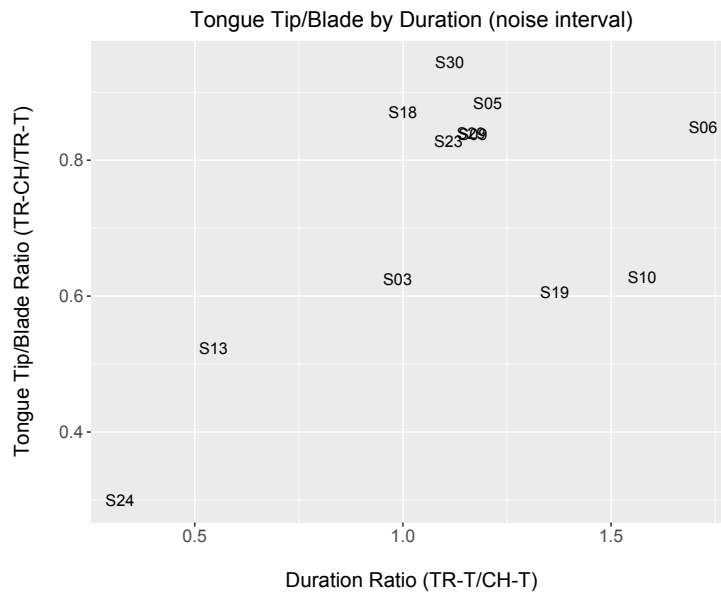
### 3.2.4 Comparing Tongue Contour and Duration Results

In Figure 3.34, tongue contour data (§ 3.2.2) are plotted against duration data (§ 3.2.3). It was reasoned that participants who were producing /tɹ/ and /dɹ/ that were acoustically more similar to [tʃ] and [dʒ] may also be exhibiting [tʃ]- and [dʒ]-like tongue contours at the tip/blade. Following Baker et al. (2011), and similar to the Affrication Ratio presented in Chapter 2, participants' mean duration scores were used to calculate a Duration Ratio in order to gauge how affricate-like participants' /tɹ/ productions were relative to their prevocalic /t/ and /tʃ/ productions. Duration Ratio was calculated by subtracting a participant's mean duration of the noise interval (following release) in /tɹ/ from the mean duration of the noise interval in /tʃ/, and dividing that value by the difference between the mean duration of the noise intervals in /tʃ/ and /t/. A ratio closer to 1 indicates that the participant's /tɹ/ duration is similar to their /tʃ/ duration (i.e., more [tʃ]-like) whereas a ratio closer to 0 indicates a /tɹ/ duration closer to prevocalic /t/ (i.e., more [t]-like). A ratio higher than 1 indicates that a participant's /tɹ/ duration is greater than their /tʃ/ duration. This approach allows for speaker-internal normalization, such that /tɹ/ duration is placed in context within the participant's acoustic space relative to /t/ and /tʃ/.

A similar measure - a Tongue Tip/Blade Ratio - was calculated for tongue curve differences at the tongue tip/blade. The absolute difference between participants' /tɹ/ and /tʃ/ curves was divided by the absolute difference between participants' /tɹ/ and /t/ curves at the mid-point of the noise interval in /ʌ/ vowel contexts. This vowel context was chosen because in front vowel contexts, there were often no significant differences between tongue contours at the tongue tip/blade, while in back vowel contexts, where the anterior portion of the tongue is articulatorily less encumbered, patterns between contours were observed. A ratio closer to 1 indicates that the distance between /tɹ/ and /tʃ/ is small, relative to the distance between

/tɹ/ and /t/ (i.e., /tɹ/ is more [tʃ]-like than [t]-like), whereas a ratio closer to 0 indicates that the distance between /tɹ/ and /tʃ/ is similar to the distance between /tɹ/ and /t/ (i.e., /tɹ/ is as [tʃ]-like as it is [t]-like)<sup>15</sup>.

As illustrated in Figure 3.34, in comparing across articulatory measures, different behaviour patterns were observed. Six participants (S05, S09, S18, S23, S29, S30) exhibited [tʃ]-like behaviour for both measures, with Duration Ratios close to 1 and small differences between /tɹ/ and /tʃ/ at the tongue tip/blade relative to /tɹ/ and /t/ (Tongue Tip/Blade Ratios closest to 0). One participant (S06) exhibited similarly small differences between /tɹ/ and /tʃ/ at the tongue tip/blade, but showed greater duration. The remaining five participants (S03, S10, S13, S19, S24) showed more variation across both measures. The two participants who showed the shortest (least /tʃ/-like) duration (S13, S24), also exhibited the least amount of similarity between /tɹ/ and /tʃ/ at the tongue tip/blade.



**Figure 3.34:** All participants: Tongue Tip/Blade Ratio (TR-CH/TR-T) by Duration Ratio (TR-T/CH-T).

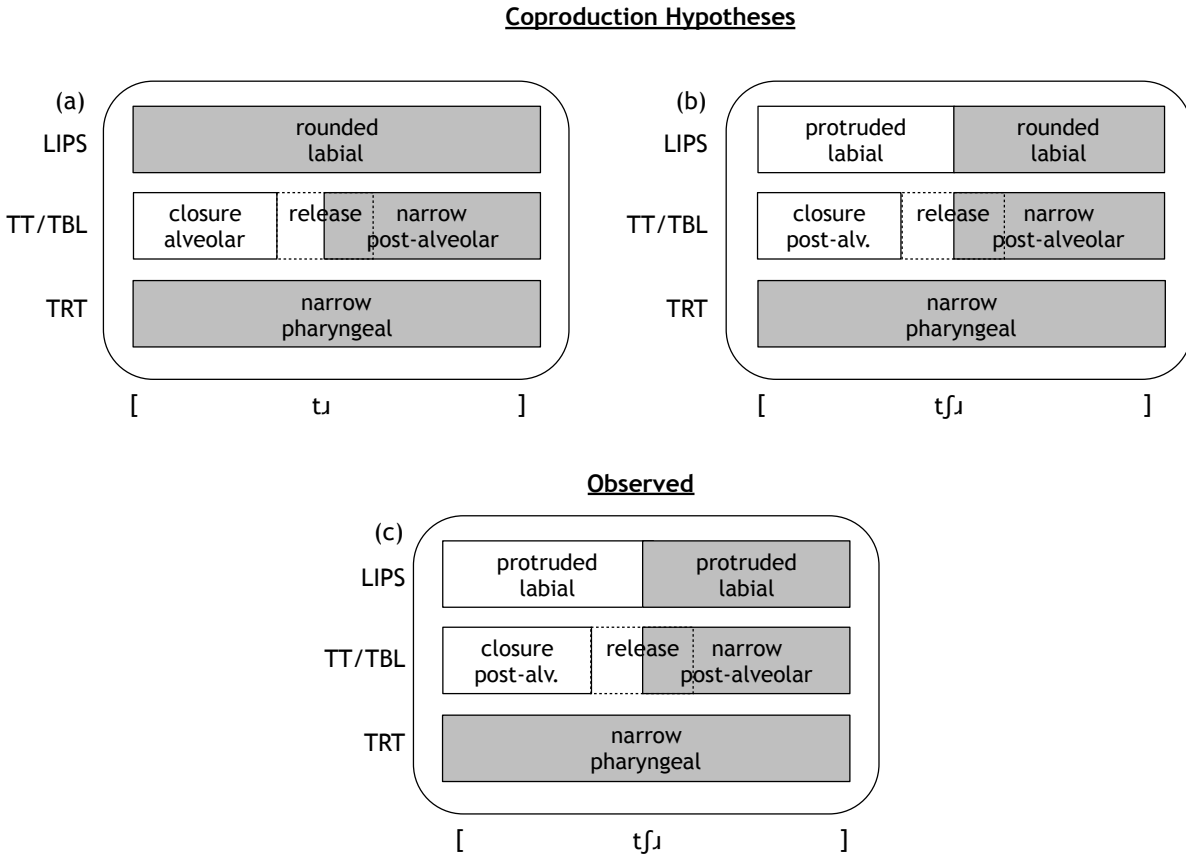
<sup>15</sup>There were no speakers who produced /tɹ/ contours closer to /t/ than /tʃ/ by the mid-point of the noise interval, but had there been, ratio values would be < 0.

### 3.3 Discussion

In the gestural score diagram depicted in Figure 3.35 (a), presented again below, it was hypothesized that /tɪ/ and /dɪ/ sequences, if coproduced, would show early engagement of the lips (LIPS) and tongue root (TRT) in anticipation of the /ɪ/ as early as at the start of the closure interval for the coronal stop. It was reasoned that nothing prohibits early engagement of these articulators, because coronal stop gestures only involve the tongue tip/blade (TT/TBL) articulatory tier. In the gestural score diagram depicted in Figure 3.35 (b), it was hypothesized that if participants were producing more [tʃ]-like /t/s before /ɪ/, we would see lip protrusion, not rounding, during the closure interval, along with more posterior ([tʃ]-like) tongue tip/blade postures, and a retracted tongue root, because nothing prohibits early engagement of this articulatory tier (TRT). Both coproduction hypotheses (a, b) predicted that articulation of the /ɪ/ would begin as early as the mid-point of the noise interval, since the tongue becomes articulatorily less encumbered at the tongue tip/blade (TT/TBL) following release of the closure.

Figure 3.35 presents the coproduction hypotheses (a and b), as well as a revised gestural score diagram (c), based on the observed data reported in the sections above. As illustrated in (c), and consistent with hypothesis (b), participants generally showed lip protrusion as early as the start of the closure interval in /tɪ/ and /dɪ/, exhibiting similar lip postures to those observed in prevocalic /tʃ/ and /dʒ/. Importantly, rounded postures were not observed at this time point (as was hypothesized in (a)), with all participants producing protruded lip shapes that were distinctly different from their more rounded word-initial /ɪ/ postures. Lip protrusion persisted through to the mid-point of the noise interval following release - a result that is not consistent with either hypothesis, both of which predicted rounded (word-initial-[ɪ]-like) lip postures by this point in time. Overall, lip results are compatible with those reported by Mielke, Smith, and Fox (2017b), who used linear discriminant analysis

(LDA) of pixel intensities in ultrasound images to confirm that the lip shape of /ʃ/ is different from the rounding of word-initial /ɪ/.



**Figure 3.35:** Top row (a, b): Coproduction hypotheses for [tɪ] (left) and [tʃɪ] (right) sequences, originally presented in Figure 3.1. Bottom row (c): Revised gestural score diagram, based on observed data.<sup>16</sup>

Also illustrated in (c) and consistent with hypothesis (b), participants frequently showed /tɪ/ and /dɪ/ tongue tip/blade postures at the mid-point of the closure interval that closely resembled prevocalic /tʃ/ and /dʒ/, and were significantly different from prevocalic /t/ and /d/. This was particularly apparent in back vowel and lax front vowel contexts, while in /i/ vowel contexts, where the anterior portion of the tongue is more actively engaged, there were typically no significant differences

<sup>16</sup>For lack of a better way to represent a post-alveolar affricate (followed by [ɪ]) that is [tʃ]-like, Hypothesis (c) represents these coproduced gestures using the transcription [tʃɪ].

between contours, or */tɹ/* and */dɹ/* showed lower tongue tip/blade postures relative to the other two contours. Importantly, with the exception of S24, participants were not exhibiting */tɹ/* and */dɹ/* postures that were articulatorily similar to prevocalic */t/* and */d/* at the mid-point of closure, as predicted in hypothesis (a), except in cases where there were no significant differences between any of the contours at the tip/blade.

This pattern persisted through to the mid-point of the noise interval, with all participants (including S24) consistently producing */tɹ/* and */dɹ/* postures at the tip/blade that were significantly different from */t/* and */d/* in */ɪ ʌ ɑ/* vowel contexts. For most participants, */tɹ/* and */tʃ/* and */dɹ/* and */dʒ/* postures patterned together, though a number of participants were exhibiting three distinct tongue contours by the mid-point of the noise interval. Deviating from the overall pattern, in */i/* vowel contexts, participants still showed little or no differences at the tongue tip/blade mid-noise, or displayed lower */tɹ/* and */dɹ/* postures, relative to the other contours. This is consistent with the notion that the anterior portion of the tongue is actively engaged during anticipatory coarticulation of a tense high front vowel like */i/*, which may result in higher tongue postures for all contours, or similar tongue heights for prevocalic */t d tʃ dʒ/*, relative to */tɹ/* and */dɹ/*, where [ɹ] blocks the coarticulatory effect of [i].

Hypotheses (a) and (b) predicted that participants would produce [r] tongue shapes at the tip/blade during the interval following release. Olive, Greenwood, and Coleman (1993:280) have argued:

For all three voiceless stops, the center of the */r/* occurs during the aspiration region and before the onset of voicing. The production of */r/* during the aspiration is possible because after the closure is released, the tongue is free to retroflex.

While the authors reference retroflex tongue shapes, post-coronal contexts like */tɹ/*

and /dɪ/ are more conducive to bunched /ɪ/ postures (Mielke et al., 2010; Magloughlin, 2016), which participants produced almost exclusively (S23 was the exception, frequently producing retroflex /ɪ/s in these contexts). Hypotheses (a) and (b) are based on the assumption that the tongue is also ‘free to bunch’ after the closure is released. Given that participants were producing [tʃ]-like and [dʒ]-like /tɪ/ and /dɪ/ contours at the mid-point of the closure interval, and given that /ʃ/, /ʒ/, and bunched /ɪ/ involve similar articulations (tongue blade raising with post-alveolar constriction), the patterning together of /tɪ/ and /tʃ/ and /dɪ/ and /dʒ/ at the mid-point of the noise interval is consistent with the hypotheses, and reflected in Figure 3.35 (c). Distinguishing between [tʃ], [dʒ], and [ɪ] contours was facilitated in instances where participants were exhibiting either concavity at the tongue dorsum (e.g., Figure 3.29) – a pattern observed for most participants, and associated with the production of a bunched /ɪ/ – or a retroflex tongue posture, as was observed for S23. For participants producing three distinct tongue contours at this time point (e.g., Figure 3.23), distinct bunched /ɪ/ postures were particularly apparent.

Also consistent with hypotheses (a) and (b), the gestural score in Figure 3.35 (c) reflects that participants typically exhibited retracted tongue root postures in /tɪ/ and /dɪ/ sequences relative to the other two contours as early as at the mid-point of the closure interval, as well as at the mid-point of the noise interval following release. Tongue root retraction in /tɪ/ and /dɪ/ relative to prevocalic /t d tʃ dʒ/ was most extreme in front vowel contexts, which is in keeping with the tongue root advancement we expect to see during production of high front vowels as a result of tongue body raising, and the tongue root retraction we anticipate during production of [ɪ] due to its pharyngeal constriction. In back vowel contexts (particularly /ɑ/), participants’ tongue roots in prevocalic /t/ and /d/ were frequently as retracted or sometimes more retracted than /tɪ/ and /dɪ/.

Although the noise interval following release was depicted in the hypothesized

gestural score diagrams (a and b), there were no explicit predictions about duration. Nevertheless, results from the duration data provided perhaps the most compelling evidence that /t/s and /d/s before /ɹ/ are behaving differently from /t/s and /d/s before vowels. Duration of the noise interval in /tɹ/ and /dɹ/ sequences was consistently significantly longer than in prevocalic /t/ and /d/, but similar in duration to prevocalic /tʃ/ and /dʒ/. Affricates are expected to have a longer period of frication noise than stops following release of the closure gesture (Byrd, 1993), but stops are expected to show an extended period of turbulence following release into a segment with a high tongue position (Ohala and Solé, 2010). Coronal stops before /i/ and /ɹ/ both meet the criteria for an aerodynamic-based explanation of increased duration. A closer look at coronal stops and affricates by vowel context reveals that while participants do exhibit longer noise interval duration before /i/, as anticipated, the duration differences before /ɹ/ cannot be explained by aerodynamic factors alone. Duration of the noise interval of /tɹ/ and /dɹ/ across vowel contexts is longer than that of /t/ and /d/ before /i/, but comparable to the duration of /tʃ/ and /dʒ/ across vowel contexts.

### 3.3.1 Answering Question 2

The question posed at the beginning of this chapter asked:

**Q2:** Are English /t/ and /d/ in /tɹ/ and /dɹ/ clusters articulated like prevocalic [t] and [d], like prevocalic [tʃ] and [dʒ], like neither, or like both?

Taken together, results show that /t/ and /d/ in /tɹ/ and /dɹ/ clusters are not being articulated like prevocalic [t] and [d]. Rather, they are being produced like [tʃ] and [dʒ] at the lips (LIPS) and the tongue tip/blade (TT/TBL) as early as the mid-point of the closure interval, but like [ɹ] at the tongue root (TRT), and like [tʃ] and [dʒ] at the lips (LIPS) at the mid-point of the noise interval, but like [tʃ] or [dʒ]

and [ɹ] at the tongue tip/blade (TT/TBL), and like [ɹ] at the tongue root (TRT). In other words, /tɹ/ and /dɹ/ sequences are being coproduced, and the tongue is engaged in two separate articulatory goals simultaneously: /ɹ/ at the tongue root, and post-alveolar affricate at the tongue tip/blade – a finding further supported by duration results. The revised gestural score diagram depicted in Figure 3.35 (c) represents these /tɹ/ and /dɹ/ sequences as coproduced gestures that are at once part post-alveolar affricate and part [ɹ]. Rather than simply the coarticulation of /tɹ/ and /dɹ/, Production study participants are coproducing gestures that look more like the *aftermath* of coarticulation – they have phonologized what was once a coarticulatory effect.

# Chapter 4

## Perception Study

The Perception Study was designed to explore how listeners categorize /t/s before /ɪ/. It was reasoned that if speakers are producing an affricated variant of /t/, and listeners are perceiving that affrication, listeners may categorize the variant as CH rather than T in a forced-choice spelling task. The Perception experiment consisted of a series of non-words because affricates in complex onsets are prohibited in English. As a result, English has no contrastive /tɪ/ - /tʃɪ/ sequences, meaning words produced as [t.ɪ] and [tʃ.ɪ] are both likely to be categorized as *tree* by listeners, while non-words [t.ɪ.ʌv] and [tʃ.ɪ.ʌv] may be categorized differently. The Perception Study asked the question:

**Q3:** Do English speakers categorize affricated variants of /t/ found in /tɪ/ clusters as T or CH?

### 4.1 Methodology

The experiment consisted of a forced-choice spelling task in which participants heard an audio stimulus over headphones, and were then presented with two spelling choices on the left and right sides of a computer screen. For example, a participant heard [tʌv], and was presented with: TOV (left side of screen), and CHOV

(right side of screen). The participant pressed the left arrow key to select TOV, or the right arrow key to select CHOV. The experiment was designed to capture the participant’s categorizations of affricated variants of /t/ that had been spliced from before /ɹ/, as well as their categorizations of prevocalic /t/ and /tʃ/. For experimental (non-filler) stimuli, each forced-choice pair involved two spelling options that differed only in terms of whether they were spelled with a T or a CH (e.g. TUV vs. CHUV, TRIV vs. CHRIV, etc.). The audio stimulus presented with each forced-choice spelling pair consisted of one of three initial sounds, concatenated with either a vowel-initial sequence (e.g., IV), or an [ɹ]-initial sequence (e.g., RIV):

- [tʃ] spliced from before a vowel (+ vowel-initial, or + [ɹ]-initial sequence)
- [t] spliced from before a vowel (+ vowel-initial, or + [ɹ]-initial sequence)
- [t] spliced from before /ɹ/ (+ vowel-initial, or + [ɹ]-initial sequence)

#### 4.1.1 Experimental Design

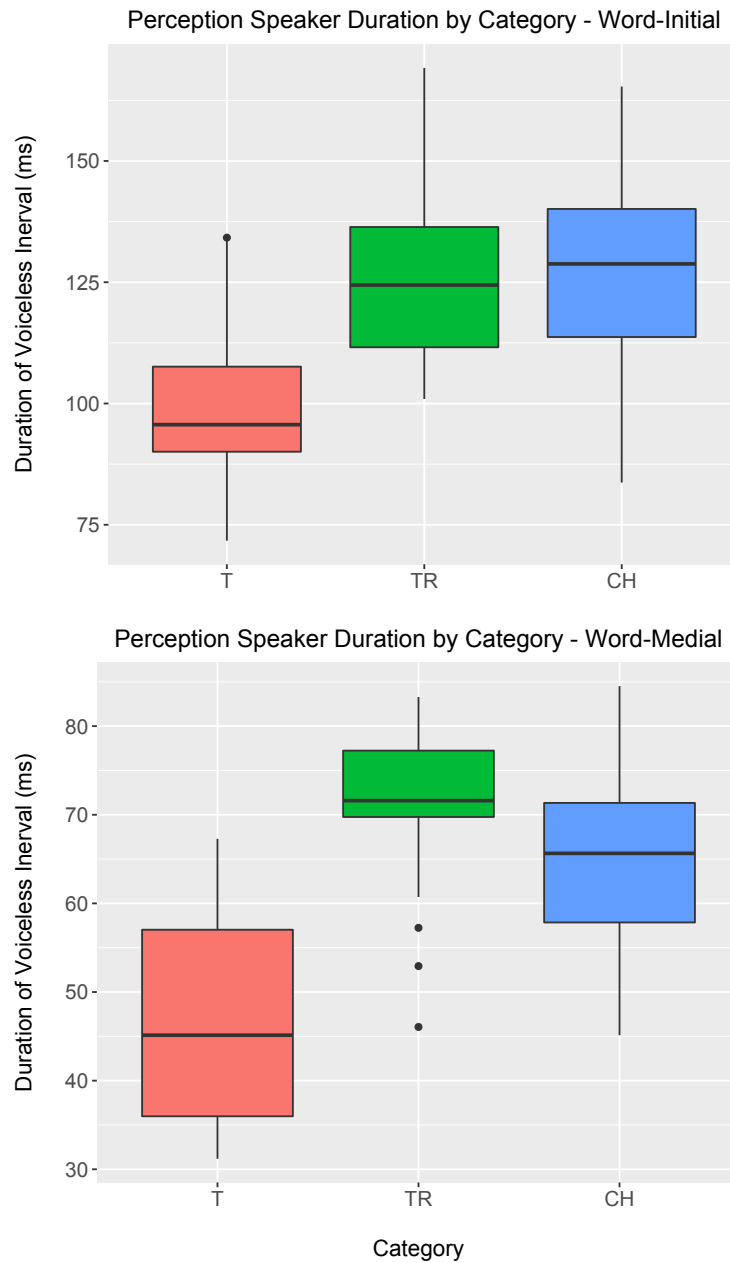
The experiment was created using a graphical open-source experiment builder called Open Sesame v3.0.7 (Mathôt et al., 2012), running PsychoPy (Peirce, 2007) on the back-end, and Python v2.7 (Van Rossum et al., 2007). A total of 92 audio stimuli (72 experimental, 20 fillers) were played for each participant over headphones, in four pseudorandomized trials ( $92 \times 4 = 368$ ). Forced-choice spelling pairs were counterbalanced (i.e. each spelling choice appeared on the left side of the computer screen in half of the trials, and on the right side in the other half). An additional 8 audio stimuli (8 fillers, in two pseudorandomized trials,  $8 \times 2 = 16$ ) were played for each participant during a Practice Phase, prior to beginning the Experimental Phase.

### 4.1.2 Stimuli

Non-words for constructing the stimuli were recorded in a sound-attenuated booth in the Phonology Lab at North Carolina State University. The recordings were produced naturally by one 24-year-old female speaker who was born and raised in Manteo, North Carolina, went to high school in Wallingford, Connecticut (age 14 to 18), and was a student at North Carolina State University in Raleigh at the time of recording. While I was not aware of any differences between her speech and the speech of Raleigh native speakers of a similar age, measures were taken to establish similarity between her affricated variants and those of the participants (as described below), most of whom were born and raised in North Carolina. The speaker was provided with a list of stimuli and filler words and asked to repeat each of the words three times (Figure 4.1), using a clear and consistent voice, with no overlap between utterances. She was given an opportunity to practice prior to recording, and was provided with feedback on pace, volume, stress of multi-syllable non-words, and vowel pronunciation during practice. The speaker was not given any details about the overall experimental design or the reason for the study. Audio was collected using a Shure BETA 53 condenser microphone worn by the speaker during recording, positioned approximately 3cm from the corner of her mouth. Recording was through a SoundDevices USBPre 2.0 preamplifier in Audacity. Figure 4.2 presents duration information (in milliseconds) for the noise interval (from release of closure to onset of voicing) of the speaker's productions of /t/ and /tʃ/ before vowels (T, CH), and /t/ before /ɹ/ (TR), in both word-initial (top) and word-medial (bottom) positions. Duration patterns were consistent with those typically exhibited by the participants in the Production experiment (§ 3.2.3), showing significantly greater duration of the noise interval in /t/ before /ɹ/ and /tʃ/ before vowels relative to /t/ before vowels.

|    |      |  |                    |    |            |  |
|----|------|--|--------------------|----|------------|--|
| 1  | TRIV |  | (rhymes with GIVE) | 28 | VETRUZZEN  |  |
|    | TRIV |  |                    |    | VETRUZZEN  |  |
|    | TRIV |  |                    |    | VETRUZZEN  |  |
| 2  | TIV  |  |                    | 29 | VETUZZEN   |  |
|    | TIV  |  |                    |    | VETUZZEN   |  |
|    | TIV  |  |                    |    | VETUZZEN   |  |
| 3  | CHIV |  |                    | 30 | VECHUZZEN  |  |
|    | CHIV |  |                    |    | VECHUZZEN  |  |
|    | CHIV |  |                    |    | VECHUZZEN  |  |
| 4  | TRIB |  |                    | 31 | METROFFEN  |  |
|    | TRIB |  |                    |    | METROFFEN  |  |
|    | TRIB |  |                    |    | METROFFEN  |  |
| 5  | TIB  |  |                    | 32 | METOFFEN   |  |
|    | TIB  |  |                    |    | METOFFEN   |  |
|    | TIB  |  |                    |    | METOFFEN   |  |
| 6  | CHIB |  |                    | 33 | MECHOFFEN  |  |
|    | CHIB |  |                    |    | MECHOFFEN  |  |
|    | CHIB |  |                    |    | MECHOFFEN  |  |
| 7  | TRUV |  | (rhymes with LOVE) | 34 | VETROBBEN  |  |
|    | TRUV |  |                    |    | VETROBBEN  |  |
|    | TRUV |  |                    |    | VETROBBEN  |  |
| 8  | TUV  |  |                    | 35 | VETOB BEN  |  |
|    | TUV  |  |                    |    | VETOB BEN  |  |
|    | TUV  |  |                    |    | VETOB BEN  |  |
| 9  | CHUV |  |                    | 36 | VECHOB BEN |  |
|    | CHUV |  |                    |    | VECHOB BEN |  |
|    | CHUV |  |                    |    | VECHOB BEN |  |
| 10 | TRUD |  | (rhymes with MUD)  | 37 | TLIF       |  |
|    | TRUD |  |                    |    | TLIF       |  |
|    | TRUD |  |                    |    | TLIF       |  |
| 11 | TUD  |  |                    | 38 | TIF        |  |
|    | TUD  |  |                    |    | TIF        |  |

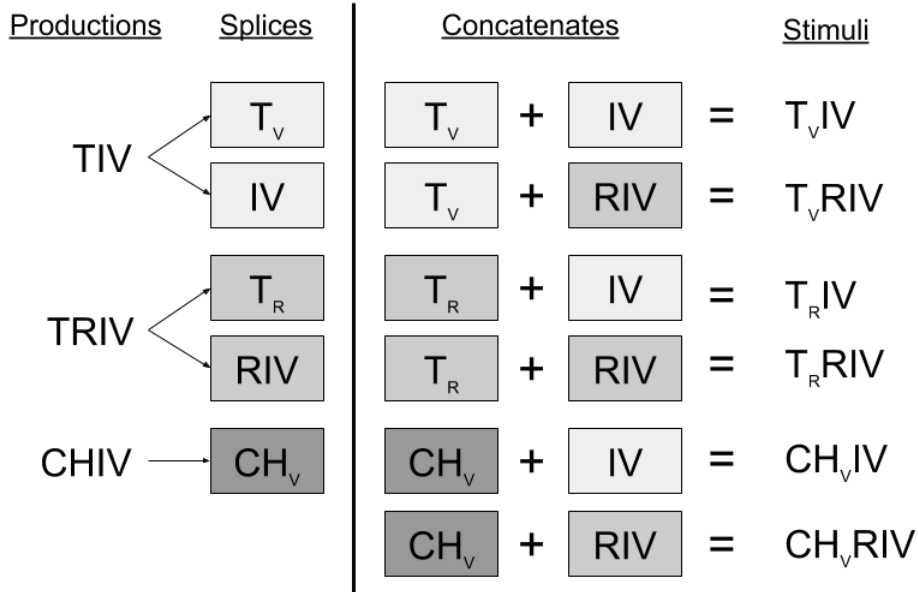
**Figure 4.1:** Excerpt from the list of non-words given to the speaker during recording of Perception stimuli for splicing. Each non-word was repeated three times by the speaker, who was instructed to use a clear and consistent voice.



**Figure 4.2:** Perception Experiment Stimuli: Duration of the noise interval (ms), from release of closure to onset of voicing, in speaker's /t/ before vowels (T), /t/ before /ɪ/ (TR), and /tʃ/ before vowels (CH), in word-initial (top) and word-medial position (bottom).

#### 4.1.2.1 Construction of Stimuli

Experimental stimuli consisted of non-words with /t/, and /tʃ/ in initial and medial (onset of the second syllable) position. As illustrated in Figure 4.3, stimuli were constructed using spliced portions of the items recorded by the speaker that were then concatenated to form stimuli. For example, recordings of TIV, TRIV, and CHIV were spliced to obtain prevocalic [t] ( $T_V$ ), the affricated variant of /t/<sup>1</sup> before /ɪ/ ( $T_R$ ), and prevocalic [tʃ] ( $CH_V$ ). Vowel-initial and [ɪ]-initial sequences were spliced from non-words (e.g., IV from TIV, RIV from TRIV, etc.), and each concatenated with the three stimuli pairs in the set. In other words, IV spliced from TIV was used to create  $T_VIV$ ,  $T_RIV$ , and  $CH_VIV$ , while RIV spliced from TRIV was used to create  $T_VRIV$ ,  $T_RRIV$ , and  $CH_VRIV$ . A complete list of experimental stimuli is presented in Table 4.1.



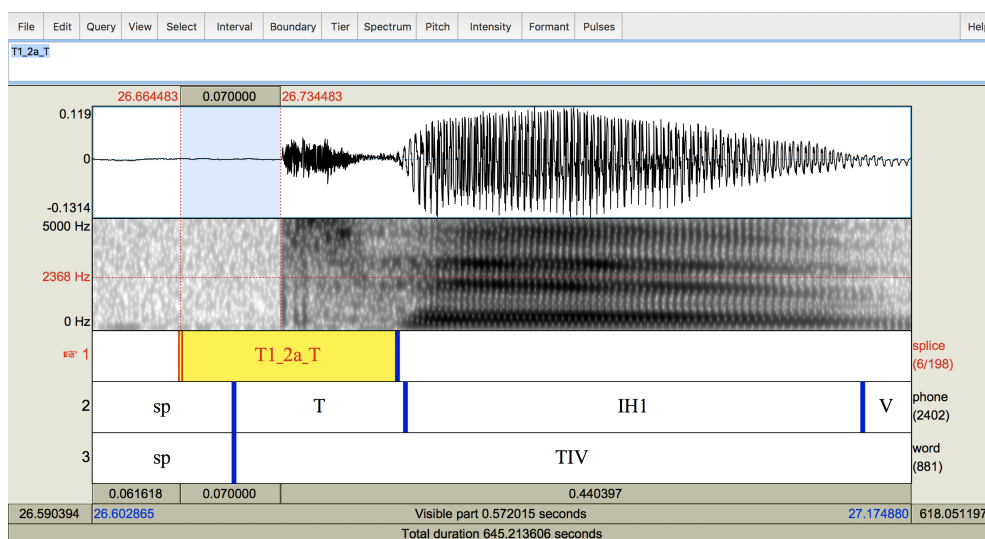
**Figure 4.3:** Construction of Perception experiment stimuli

<sup>1</sup>Throughout this chapter, I will refer to the affricated variant of /t/ before /ɪ/ in order to distinguish it from prevocalic /t/, based on the assumption that it is not phonetically [t] (see Chapter 3), but may be perceived as such.

| Vowel Context | Category                 |                          |
|---------------|--------------------------|--------------------------|
|               | T <sub>V</sub> V tokens  | T <sub>V</sub> R tokens  |
| [ɪ]           | T <sub>V</sub> IV        | T <sub>V</sub> RIV       |
|               | T <sub>V</sub> IB        | T <sub>V</sub> RIB       |
|               | MET <sub>V</sub> IFFEN   | MET <sub>V</sub> RIFFEN  |
|               | VET <sub>V</sub> IBBEN   | VET <sub>V</sub> RIBBEN  |
| [ʌ]           | T <sub>V</sub> UV        | T <sub>V</sub> RUV       |
|               | T <sub>V</sub> UD        | T <sub>V</sub> RUD       |
|               | MET <sub>V</sub> UPPEN   | MET <sub>V</sub> RUPPEN  |
|               | VET <sub>V</sub> UZZEN   | VET <sub>V</sub> RUZZEN  |
| [ɑ]           | T <sub>V</sub> OV        | T <sub>V</sub> ROV       |
|               | T <sub>V</sub> OB        | T <sub>V</sub> ROB       |
|               | MET <sub>V</sub> OFFEN   | MET <sub>V</sub> ROFFEN  |
|               | VET <sub>V</sub> OBEN    | VET <sub>V</sub> ROBBEN  |
| Vowel Context | T <sub>R</sub> V tokens  | T <sub>R</sub> R tokens  |
| [ɪ]           | T <sub>R</sub> IV        | T <sub>R</sub> RIV       |
|               | T <sub>R</sub> IB        | T <sub>R</sub> RIB       |
|               | MET <sub>R</sub> IFFEN   | MET <sub>R</sub> RIFFEN  |
|               | VET <sub>R</sub> IBBEN   | VET <sub>R</sub> RIBBEN  |
| [ʌ]           | T <sub>R</sub> UV        | T <sub>R</sub> RUV       |
|               | T <sub>R</sub> UD        | T <sub>R</sub> RUD       |
|               | MET <sub>R</sub> UPPEN   | MET <sub>R</sub> RUPPEN  |
|               | VET <sub>R</sub> UZZEN   | VET <sub>R</sub> RUZZEN  |
| [ɑ]           | T <sub>R</sub> OV        | T <sub>R</sub> ROV       |
|               | T <sub>R</sub> OB        | T <sub>R</sub> ROB       |
|               | MET <sub>R</sub> OFFEN   | MET <sub>R</sub> ROFFEN  |
|               | VET <sub>R</sub> OBEN    | VET <sub>R</sub> ROBBEN  |
| Vowel Context | CH <sub>V</sub> V tokens | CH <sub>V</sub> R tokens |
| [ɪ]           | CH <sub>V</sub> IV       | CH <sub>V</sub> RIV      |
|               | CH <sub>V</sub> IB       | CH <sub>V</sub> RIB      |
|               | MECH <sub>V</sub> IFFEN  | MECH <sub>V</sub> RIFFEN |
|               | VECH <sub>V</sub> IBBEN  | VECH <sub>V</sub> RIBBEN |
| [ʌ]           | CH <sub>V</sub> UV       | CH <sub>V</sub> RUV      |
|               | CH <sub>V</sub> UD       | CH <sub>V</sub> RUD      |
|               | MECH <sub>V</sub> UPPEN  | MECH <sub>V</sub> RUPPEN |
|               | VECH <sub>V</sub> UZZEN  | VECH <sub>V</sub> RUZZEN |
| [ɑ]           | CH <sub>V</sub> OV       | CH <sub>V</sub> ROV      |
|               | CH <sub>V</sub> OB       | CH <sub>V</sub> ROB      |
|               | MECH <sub>V</sub> OFFEN  | MECH <sub>V</sub> ROFFEN |
|               | VECH <sub>V</sub> OBEN   | VECH <sub>V</sub> ROBBEN |

**Table 4.1:** Perception stimuli, by vowel context and category: [t] spliced from before vowels (T<sub>V</sub>V, T<sub>V</sub>R, upper); the affricated variant of /t/ spliced from before /ɪ/ (T<sub>R</sub>V, T<sub>R</sub>R, middle); [tʃ] spliced from before vowels (CH<sub>V</sub>V, CH<sub>V</sub>R, lower).

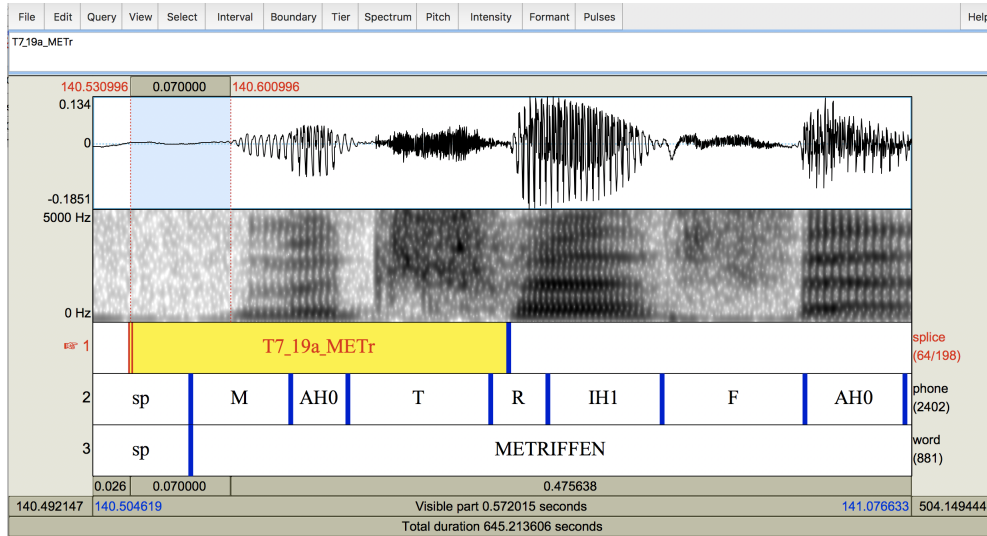
Stimuli were spliced and concatenated<sup>2</sup> in Praat (Boersma and Weenink, 2007). Stops and affricates in word-initial position before vowels and /ɪ/ were spliced from the original .wav file at 70 milliseconds before the start of the burst to the onset of voicing (Figure 4.4). Stops and affricates in word-medial position before vowels and /ɪ/ were spliced from the original file along with the preceding word-initial voiced consonant-vowel (CV) sequences at 70 milliseconds before the word-initial onset of voicing to the onset of voicing after the word-internal stop burst (Figure 4.5). Vowel-initial (e.g., IV) and [ɪ]-initial (e.g., RIV) sequences were spliced from the onset of voicing of the vowel or /ɪ/ following the stop or affricate to the end of the word plus 70 milliseconds. Splices were concatenated in Praat (as illustrated in Figure 4.3).



**Figure 4.4:** Example of a stop spliced from TIV, at -0.07 seconds from before the burst to the end of the noise interval.

Filler stimuli consisted of monosyllabic non-words that differed only in terms of word-initial voicing. Stimuli were constructed using the same methods as the experimental stimuli. For example, the speaker recorded TLIV, TIV. The [t] was

<sup>2</sup>Thank you to Erik Thomas, North Carolina State University, for assistance with stimuli creation and band-pass filtering to remove non-speech noise from two filler tokens.



**Figure 4.5:** Example of a CV+stop sequence spliced from METRIFEN, at -0.07 seconds from before the word-initial onset of voicing to the end of the noise interval for the stop.

spliced from before the vowel in TIV, and the [l]-initial sequence was spliced from the onset of voicing of the [l] in TLIV. The two splices were concatenated to form T<sub>V</sub>LIV. Experimental and filler stimuli were constructed using the same procedures in order to ensure consistency across stimuli, and so that participants were not provided with any cues as to which stimuli were experimental, and which were fillers.

#### 4.1.2.2 Experimental Trials

A total of 3888 audio trials were categorized by 27 participants in each Position (initial and medial), including 648 audio trials in each Category:  $6 \times 648 = 3888$ . (By participant:  $3888 / 27$  participants = 144 trials per participant, and 24 trials in each Category). A total of 1296 audio trials were categorized by 27 participants in each of three Vowel Contexts (/ɪ ʌ ɑ/), including a total of 216 audio trials in each Category:  $6 \times 216 = 1296$ ;  $1296 \times 3 = 3888$ . (By participant:  $1296 / 27$  participants = 48 trials per participant in each Vowel Context, and 8 in each Category.)

| Vowel Context | Filler Spelling Pairs  |
|---------------|--|
| [ɪ]           | DLIV ~ TLIV<br>DLIF ~ TLIF                                   |
| [ʌ]           | GNUB ~ KNUB<br>GNUP ~ KNUP<br>VNUD ~ FNUD<br>VNUT ~ FNUT     |
| [ɑ]           | BNOG ~ PNOG<br>BNOCK ~ PNOCK                                 |
| [u]           | VOOB ~ FOOB<br>VOOP ~ FOOP<br>GLOOB ~ KLOOB<br>GLOOP ~ KLOOP |

**Table 4.2:** Spliced filler stimuli spelling pairs with word-initial contrast, by vowel context. Filler stimuli were designed to convey to the participant that sequences of sounds prohibited in English phonotactics (e.g., onset /bn/), were acceptable in this language.

### 4.1.3 Participants

Participants for the Perception experiment were 32 English speakers, recruited from the North Carolina State University community. Five participants were excluded from analysis: S01 did not complete the language background questionnaire, S14 exhibited reading difficulties at various points during her visit, and S08, S12, and S16 each reported an L1 other than English. The 27 participants included in the analysis were 16 females and 11 males, ranging in age from 19 to 60, all born in the United States (Table 4.3). A subset of these participants were included in the Production Study analysis, presented in Chapter 3, above.

### 4.1.4 Procedures

For the Perception experiment, participants moved to an adjacent sound-attenuated booth in the same lab after completing the Production experiment. The experiments were conducted in this order to eliminate the possibility that stimuli heard during the Perception experiment could influence production. Participants were seated in

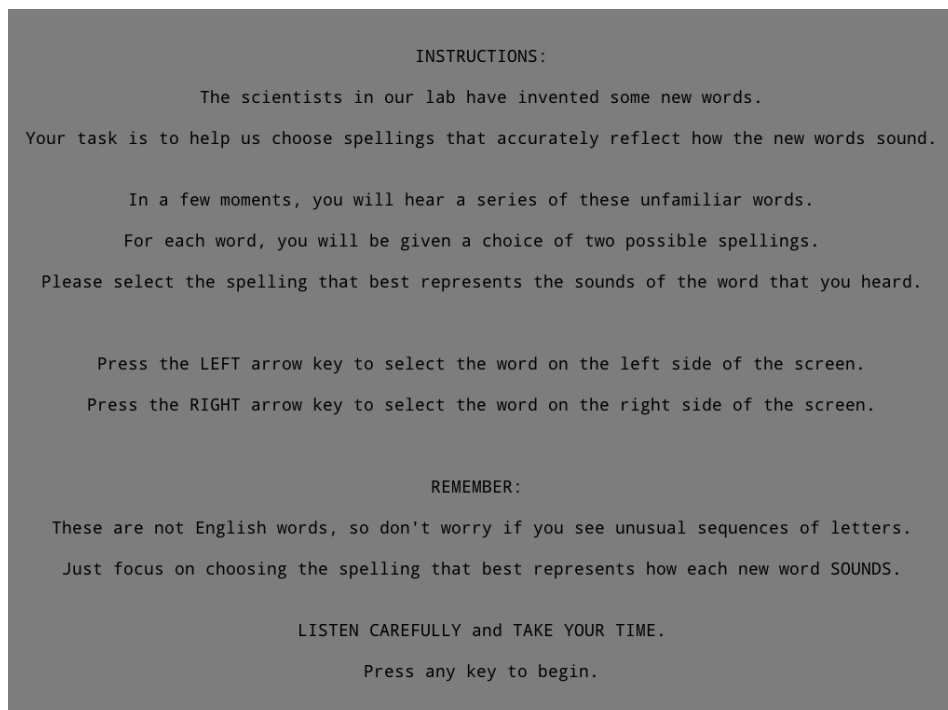
| Participant | Sex    | Birth Year | Birth Place    |
|-------------|--------|------------|----------------|
| S02         | Male   | 1990       | North Carolina |
| S03         | Female | 1996       | North Carolina |
| S04         | Female | 1994       | North Carolina |
| S05         | Female | 1995       | North Carolina |
| S06         | Female | 1996       | Wisconsin      |
| S07         | Female | 1956       | North Carolina |
| S09         | Male   | 1995       | Delaware       |
| S10         | Female | 1995       | North Carolina |
| S11         | Female | 1991       | North Carolina |
| S13         | Female | 1997       | North Carolina |
| S15         | Male   | 1988       | North Carolina |
| S17         | Female | 1993       | New Jersey     |
| S18         | Male   | 1995       | North Carolina |
| S19         | Female | 1996       | North Carolina |
| S20         | Female | 1997       | North Carolina |
| S21         | Female | 1970       | New York       |
| S22         | Male   | 1996       | North Carolina |
| S23         | Male   | 1996       | North Carolina |
| S24         | Female | 1985       | North Carolina |
| S25         | Male   | 1993       | Georgia        |
| S26         | Male   | 1979       | Tennessee      |
| S27         | Male   | 1994       | Virginia       |
| S28         | Male   | 1994       | North Carolina |
| S29         | Female | 1994       | North Carolina |
| S30         | Female | 1991       | North Carolina |
| S31         | Male   | 1996       | North Carolina |
| S32         | Female | 1964       | North Carolina |

**Table 4.3:** Perception Study participant details

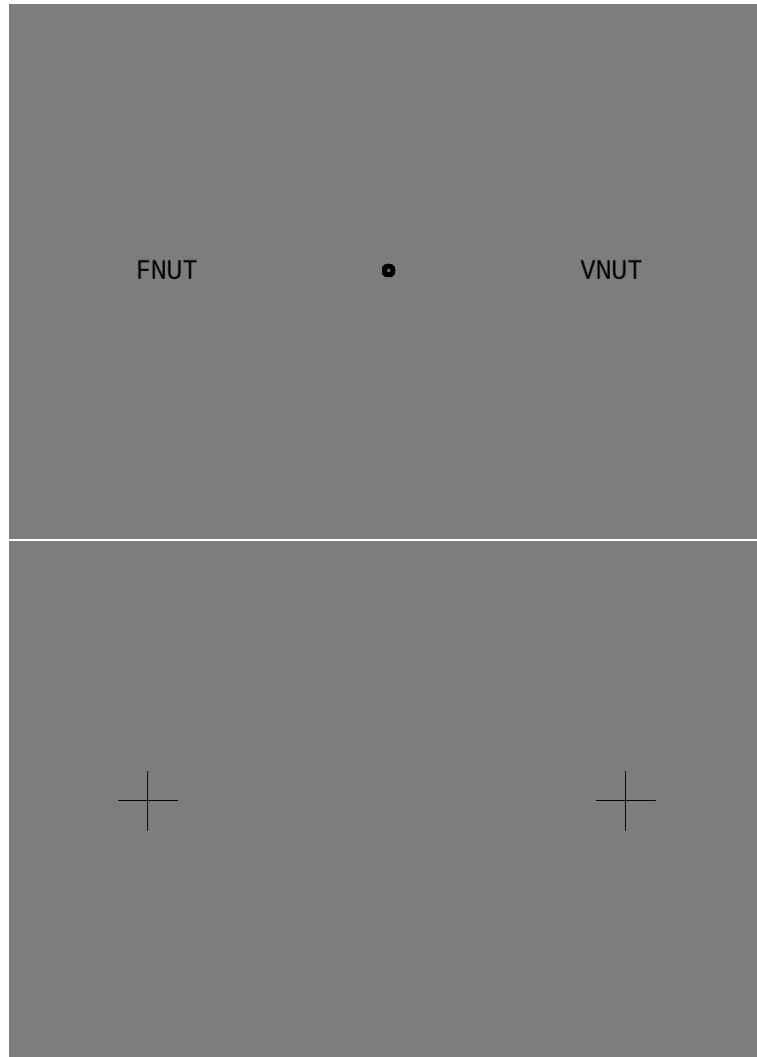
a chair in front of a computer screen and given a pair of headphones for use during the experiment, with the volume set to a comfortable listening level. Participants were told that they would hear a series of new and unfamiliar words, presented one at a time, and be asked to select one of two spellings on the screen that best reflected the sounds that they heard, using the left and right arrow keys on the keyboard. They were told that they would be presented with instructions prior to beginning a Practice Phase, which would be followed by an Experimental Phase, and that the entire process would take approximately 30 minutes. After ensuring

participants were comfortable and had no questions, the experimenter exited the booth and closed the door.

At the beginning of the experiment, participants were presented with on-screen instructions that described how the scientists in our lab had invented some new, non-English words (Figure 4.6). The first round of the experiment involved a Practice Phase where participants were presented with forced-choice spelling pairs ( $2 \times 8 = 16$ ) of monosyllabic filler words that included phonotactically restricted consonant sequences in word-initial position (e.g., FNUT vs. VNUT, Figure 4.7, top), each contrasting in terms of voicing. The use of these filler non-word pairs was intended to reinforce the idea that the new words were not English and could contain sequences of sounds that are phonotactically illicit in English. Following each key press, crosses in place of spelling choices were displayed for 500 milliseconds (Figure 4.7, bottom), prior to the next audio stimulus being played and two new spelling choices being displayed. Once finished, the participants were shown a message indicating that they had completed the Practice Phase and could move on to the Experimental Phase whenever they were ready. The message also reminded participants to select the word on the left side of the screen using the left arrow key, and the word on the right side of the screen using the right arrow key. The Experimental Phase was comprised of experimental and filler stimuli that were not heard during the Practice Phase. As illustrated in Figure 4.8, spelling pairs for experimental stimuli consisted of choices that included T or CH, before a vowel (top) or an R (bottom), in word-initial (top) or word-medial (bottom) position. Audio stimuli included [t] or [tʃ] spliced from before vowels, or the affricated variant of /t/ spliced from before /ɹ/, concatenated with a vowel-initial (e.g., IV) or [ɹ]-initial (e.g., RIV) sequence (§ 4.1.2). Participants were able to move through the Perception experiment at their own pace. Following completion of the experiment, participants filled out a short online language background questionnaire (Appendix B).



**Figure 4.6:** On-screen instructions at the beginning of the Perception experiment.



**Figure 4.7:** The practice phase of the Perception experiment consisted of forced-choice spelling pairs of non-word fillers containing sequences of sounds that are phonotactically restricted in English (top). After each selection, crosses in place of spelling choices were displayed for 500 milliseconds (bottom), prior to the next token being played and the next spelling choices being displayed.



**Figure 4.8:** Spelling pairs for experimental stimuli consisted of choices that included T or CH, before a vowel (top) or an R (bottom), in word-initial (top) or word-medial position (bottom).

## 4.2 Perception Results

The results presented here consist of participants' categorizations of audio stimuli as either T or CH, based on their selections in forced-choice spelling pairs (e.g. TROV vs. CHROV). As described above (§ 4.1.2), audio stimuli were spliced and concatenated into six categories: [tʃ] spliced from before a vowel, concatenated with a vowel-initial (CH<sub>V</sub>V) or [ɹ]-initial (CH<sub>V</sub>R) sequence, [t] spliced from before a vowel, concatenated with a vowel-initial (T<sub>V</sub>V) or [ɹ]-initial (T<sub>V</sub>R) sequence, and the affricated variant of /t/ spliced from before /ɹ/, concatenated with a vowel-initial (T<sub>R</sub>V) or [ɹ]-initial (T<sub>R</sub>R) sequence.

The experiment was designed to test whether affricated variants of /t/ would be categorized as T or CH in a forced-choice spelling task. Categorizations of T<sub>R</sub>R audio stimuli as T may not be straightforward to interpret because there are several possible reasons why a participant might choose a TR spelling over a CHR spelling (e.g. TROV over CHROV):

1. The affricated variant of /t/ sounds normal in the context that conditions the affrication (i.e., before [ɹ]) and may not be perceived as affricated;
2. /tʃɹ/ is a phonotactically restricted sequence in English, which may result in a categorization bias against an illicit cluster;
3. CHR sequences in English spellings already exist (Christmas, chrysalis, chrysanthemum, etc.), but they are pronounced [kɹ].

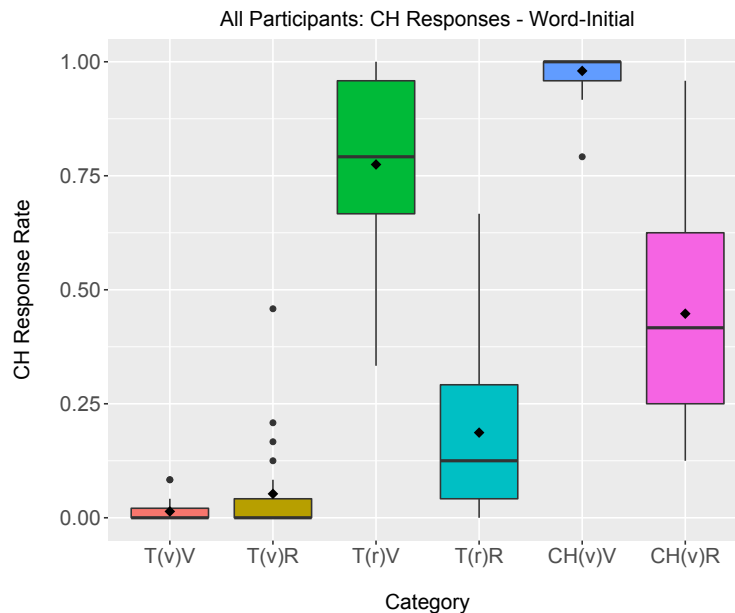
An abundance of research has shown that listeners are able to perceptually compensate for coarticulation (Mann and Repp, 1980; Whalen, 1981; Smits, 2001). Mann and Repp (1980) observed that listeners were more likely to report hearing *s* along an [s] to [ʃ] continuum when the audio token preceded [u] than when it preceded [a]. Lip-rounding lowers frequency, and [ʃ] is lower in frequency than [s]. When

participants heard tokens along the continuum, they attributed lower frequency values to lip-rounding on the vowel, rather than to the consonant. Similarly, Beddor and Krakow (1999) showed that English listeners were able to attribute nasality on a nasalized vowel to the following nasal consonant, perceiving these vowels as relatively oral. In this sense, listeners may choose TR spellings over CHR because they can attribute the aperiodic noisiness of the affricated variant of /t/ to coarticulation with the following /ɹ/. Further, Massaro and Cohen (1983) and Pitt (1998) have shown that the legality of phonotactic sequences may have an influence on how listeners categorize ambiguous phonemes. The authors found that listeners exhibited a ‘labeling bias’ in the direction of the licit sequence, suggesting that top-down influence may play a role in a participant choosing TR over CHR.

These factors, coupled with the fact that CHR spellings of actual English words are pronounced [kɹ], make the task of interpreting TR results more challenging. In an effort to (indirectly) mitigate some of these issues, the forced-choice spelling task included CH<sub>v</sub>R audio stimuli as well (as described above). It was reasoned that if participants were willing to choose CHR spellings when they heard audio stimuli with phonological [tʃ] spliced onto an [ɹ]-initial sequence, this would provide evidence that the English phonotactic bias was not a sufficient deterrent to prohibit CHR selections. Additionally, although words like [kɹɪsəlɪs] with CHR spellings *do* exist in English, it was assumed that participants would be unlikely to interpret the affricated variants they heard as [kɹ] (in the sense of Miller and Nicely, 1955), particularly since other non-word sequences with CH spellings (e.g., CHOV), where [tʃ] is straightforwardly represented by CH, were included in the experiment as well.

### 4.2.1 Overview of Results

Figure 4.9 presents participants' mean CH response rates in word-initial position across each of the six categories. (A summary of results across positions and categories is also presented in Table 4.4.) As expected, participants were able to perform the task, accurately categorizing prevocalic /t/s ( $T_VV$ ) as T, and prevocalic /tʃ/s ( $CH_VV$ ) as CH, when presented in their natural phonetic environments (i.e., before vowels). Participants categorized affricated variants of /t/ as T when presented in the environment that conditions the affrication ( $T_RR$ ), but categorized these same affricated variants of /t/ as CH when presented outside of their natural phonetic environment ( $T_RV$ ). As a group, participants showed variable behaviour in their classifications of phonological /tʃ/s, but only when heard outside of their natural phonetic environment (before [ɹ]), categorizing  $CH_VR$  stimuli as T, or CH, or a combination of both (as will be discussed below).

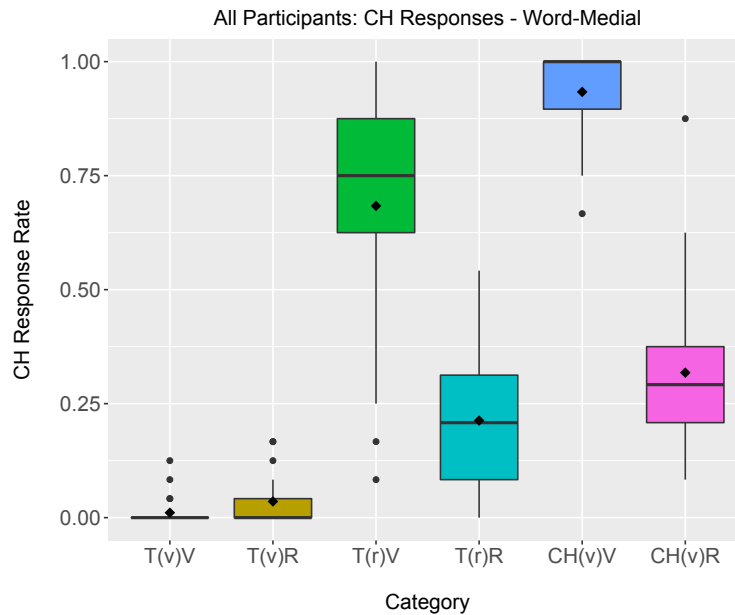


**Figure 4.9:** All Participants: CH response rate in word-initial position. (Boxes represent interquartile range of values, solid horizontal lines represent median, whiskers represent range of data points falling within 1.5 times interquartile range. Circles depict outliers.)

| Category                 | CH Response Rate |
|--------------------------|------------------|
| CH <sub>V</sub> R        | 0.45             |
| CH <sub>V</sub> R-medial | 0.32             |
| CH <sub>V</sub> V        | 0.98             |
| CH <sub>V</sub> V-medial | 0.93             |
| T <sub>R</sub> R         | 0.19             |
| T <sub>R</sub> R-medial  | 0.21             |
| T <sub>R</sub> V         | 0.77             |
| T <sub>R</sub> V-medial  | 0.68             |
| T <sub>V</sub> R         | 0.05             |
| T <sub>V</sub> R-medial  | 0.04             |
| T <sub>V</sub> V         | 0.01             |
| T <sub>V</sub> V-medial  | 0.01             |

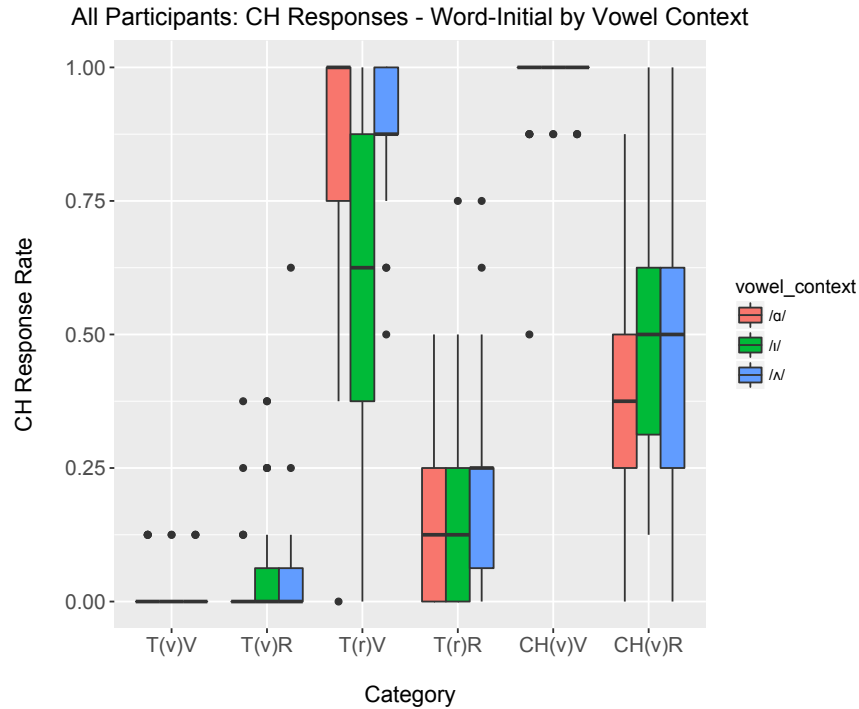
**Table 4.4:** All participants: Mean CH response rate, by category

Figure 4.10 presents participants’ mean CH response rates in word-medial position. Participants exhibited similar response rates across categories in both positions, with the exception of CH<sub>V</sub>R, which participants were more likely to categorize as T word-medially, exhibiting less variable behaviour than in word-initial position.

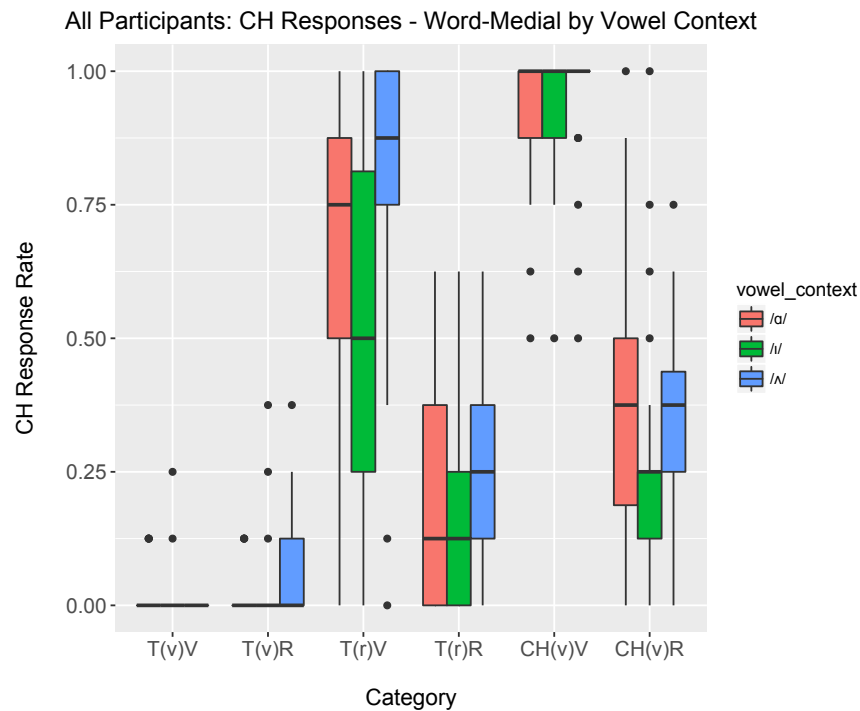


**Figure 4.10:** All participants: CH categorizations in word-medial position. (Boxes represent interquartile range of values, solid horizontal lines represent median, whiskers represent range of data points falling within 1.5 times interquartile range. Circles depict outliers.)

A look at the results by Vowel Context (Figures 4.11 and 4.12) shows participant responses do not seem to be affected by the following vowel. T<sub>R</sub>V was the exception, where participants were less likely to categorize audio stimuli in /ɪ/ vowel contexts as CH, relative to stimuli in back vowels contexts: /ʌ ɑ/, in both word-initial and word-medial positions.



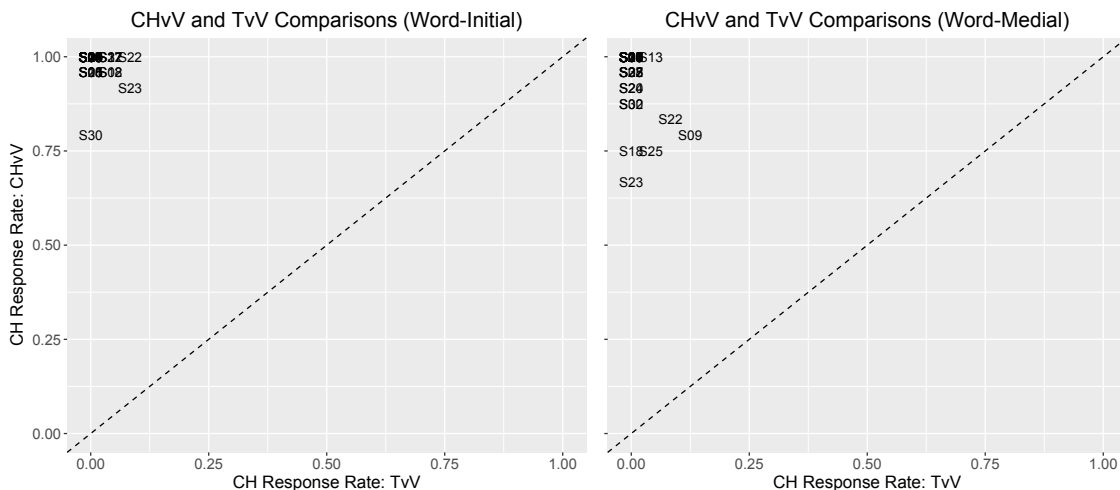
**Figure 4.11:** All participants: CH categorizations in word-initial position, by vowel context. (Boxes represent interquartile range of values, solid horizontal lines represent median, whiskers represent range of data points falling within 1.5 times interquartile range. Circles depict outliers.)



**Figure 4.12:** All participants: CH categorizations in word-medial position, by vowel context. (Boxes represent interquartile range of values, solid horizontal lines represent median, whiskers represent range of data points falling within 1.5 times interquartile range. Circles depict outliers.)

## 4.2.2 Comparing Across Categories

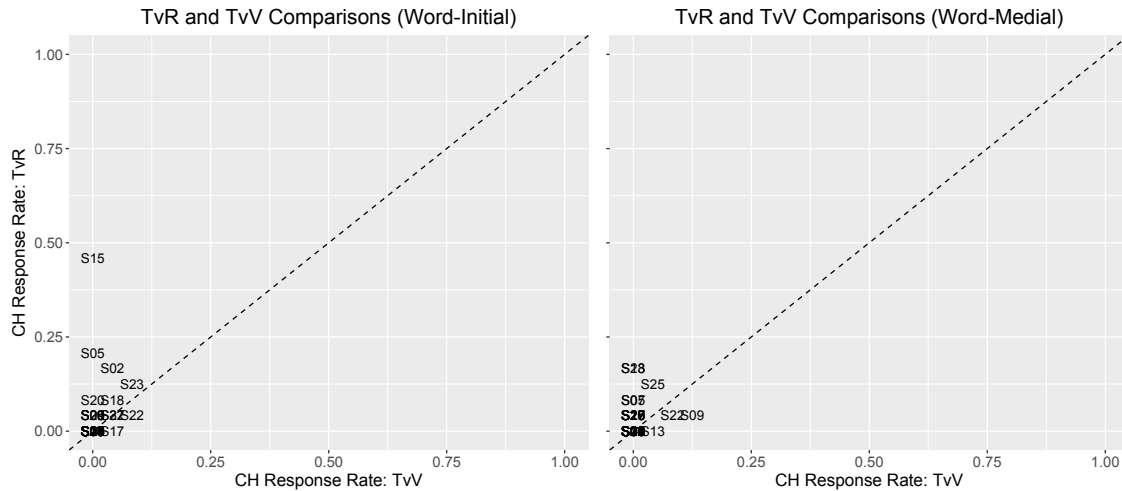
The Perception results presented in the previous section show mean values for all participants across categories and positions. The figures that follow present data points at the level of the individual participant and compare patterns across categories. It was expected that every participant would accurately categorize prevocalic [t] and [tʃ] in CH<sub>V</sub>V and T<sub>V</sub>V, because these are distinct phonemes presented in the environments where they were produced. As illustrated in Figure 4.13, participants categorized prevocalic [tʃ]s concatenated with vowel-initial sequences (CH<sub>V</sub>V) as CH (CH response rates close to 1.0, *y*-axes) in both word-initial (left) and word-medial (right) positions, and prevocalic [t]s concatenated with vowel-initial sequences (T<sub>V</sub>V) as T (CH response rates close to 0.0, *x*-axes), in both positions. In other words, [t]s and [tʃ]s spliced from before vowels (T<sub>V</sub>, CH<sub>V</sub>) were heard faithfully before vowels (e.g., T<sub>V</sub>IV, CH<sub>V</sub>IV) – confirmation that participants were able to perform the task.



**Figure 4.13:** Comparing participants' CH response rates in CH<sub>v</sub>V (*y*-axis) and T<sub>v</sub>V (*x*-axis) categories, in word-initial (left) and word-medial (right) positions.

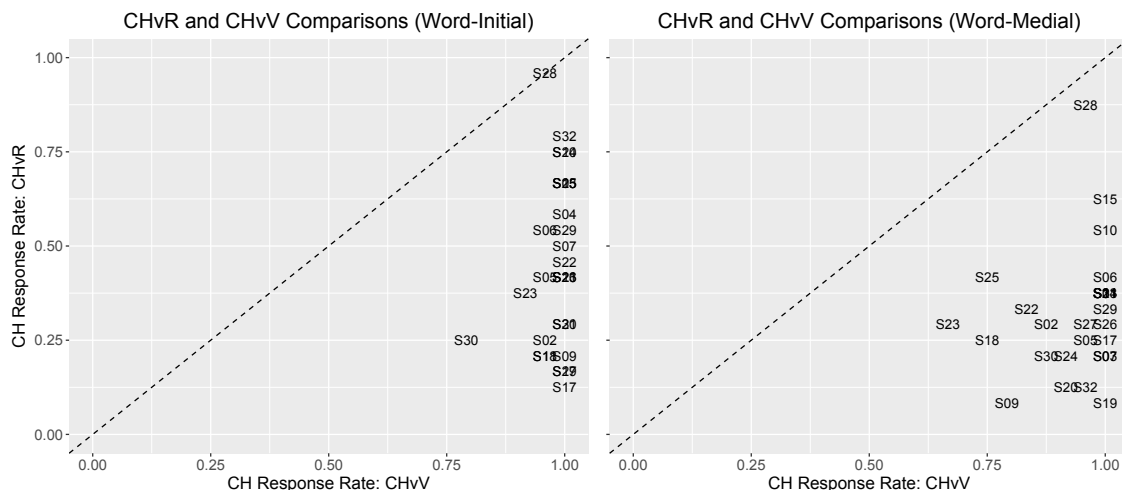
It was also anticipated that participants would categorize prevocalic [t]s that were spliced onto non-matching [ɹ]-initial sequences as T. As shown in Figure 4.14,

participants' categorizations of prevocalic [t] as T were consistent across T<sub>V</sub>R and T<sub>V</sub>V categories (CH response rates close to 0.0, *x*- and *y*-axes), in both positions. [t]s spliced from before vowels (T<sub>V</sub>) were heard faithfully, regardless of whether they were concatenated with vowel-initial or [ɹ]-initial sequences.



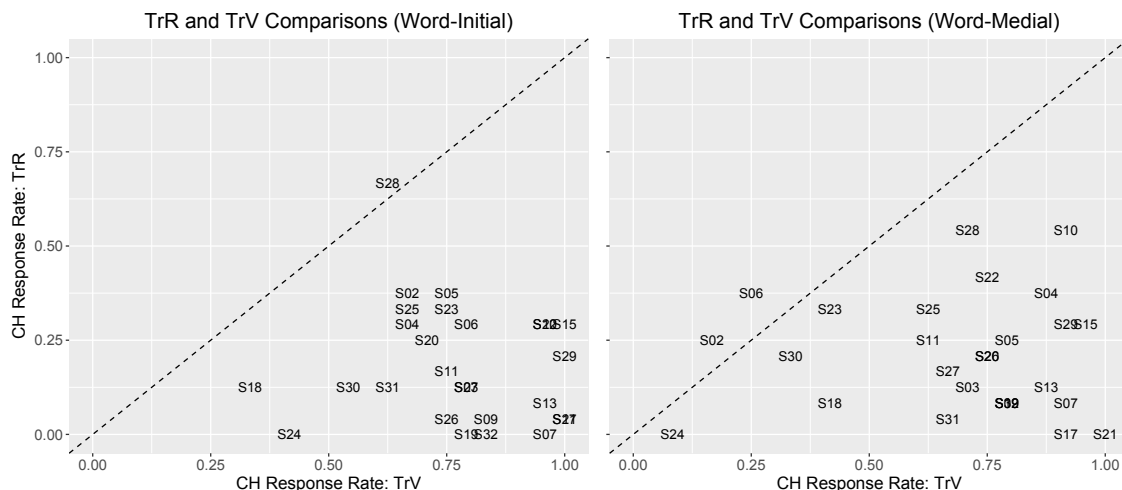
**Figure 4.14:** Comparing participants' CH response rates in TvR (*y*-axis) and TvV (*x*-axis) categories, in word-initial (left) and word-medial (right) positions.

Figure 4.15 compares participants' CH response rates in CH<sub>V</sub>R and CH<sub>V</sub>V categories. While prevocalic [tʃ]s concatenated with vowel-initial sequences (CH<sub>V</sub>V) were consistently categorized as CH in both positions, prevocalic [tʃ]s concatenated with [ɹ]-initial sequences (CH<sub>V</sub>R) were categorized as T and CH in word-initial position (left), and more consistently as T in word-medial position (right), with some inter- and intra-speaker variation in both positions (participants with high (close to 1.0), or low (close to 0.0) CH response rates, and participants with rates in between (close to 0.5)). As illustrated clearly here and in contrast with T<sub>V</sub>, participants' categorization of CH<sub>V</sub> (i.e., phonological affricates) was strongly influenced by context.



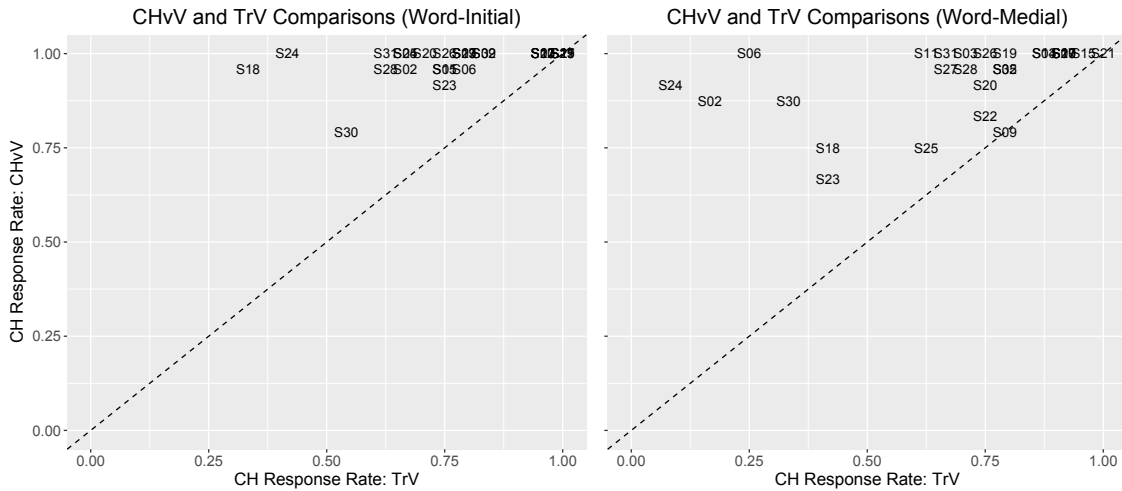
**Figure 4.15:** Comparing participants' CH response rates in CHvR ( $y$ -axis) and CHvV ( $x$ -axis) categories, in word-initial (left) and word-medial (right) positions.

Figure 4.16 compares participants' CH response rates in  $T_{R,R}$  and  $T_{R,V}$  categories. In both word-initial (left) and word-medial (right) positions, participants generally categorized affricated variants of /t/ concatenated with [ɹ]-initial sequences ( $T_{R,R}$ ) as T, but affricated variants of /t/ concatenated with vowel-initial sequences ( $T_{R,V}$ ) as CH. Results were more dispersed in  $T_{R,V}$  categories, with inter- and intra-speaker variation observed across positions. Both figures (4.15 and 4.16) show that only phonological affricates ( $CH_V$ ) were ever heard consistently as [tʃ] before [ɹ], whereas both phonological affricates ( $CH_V$ ) and affricated variants of /t/ ( $T_R$ ) were heard as [tʃ] before vowels.

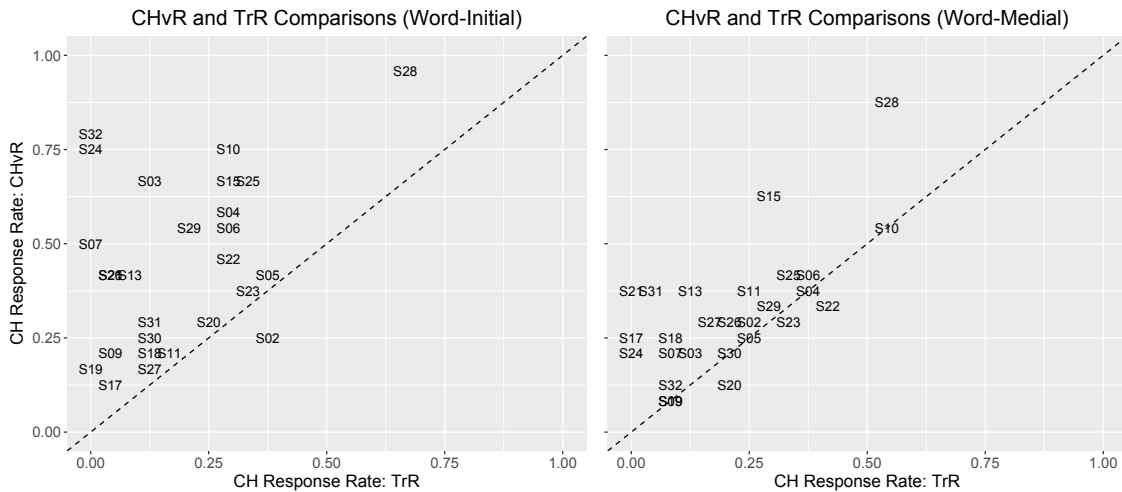


**Figure 4.16:** Comparing participants' CH response rates in TrR ( $y$ -axis) and TrV ( $x$ -axis) categories, in word-initial (left) and word-medial (right) positions.

Finally, the following figures compare phonological affricates ( $CH_V$ ) with affricated variants of /t/ ( $T_R$ ) before both vowel-initial (Figure 4.17) and [ɹ]-initial (Figure 4.18) sequences. As illustrated in Figure 4.17, before vowels, participants typically categorized both  $CH_V$  and  $T_R$  as CH, though several participants categorized  $T_R$  as T more frequently. In other words, before vowels, affricated variants of /t/ concatenated with vowels were generally perceived more like phonological affricates. In contrast, before [ɹ] (Figure 4.18), participants categorized phonological affricates ( $CH_V$ ) as T or CH, though with less variation (i.e., more T categorizations) in word-medial position, and affricated variants of /t/ ( $T_R$ ) as T. Importantly, although more perceptual variation overall was observed with phonological [tʃ] before [ɹ], at least for some participants, affricated variants of /t/ and phonological /tʃ/ were both perceived as T in pre-[ɹ] environments. Taken together, these last figures demonstrate the influence of context and position on the categorization of these sounds.



**Figure 4.17:** Comparing participants' CH response rates in CHvV (*y*-axis) and TrV (*x*-axis) categories, in word-initial (left) and word-medial (right) positions.



**Figure 4.18:** Comparing participants' CH response rates in CHvR (*y*-axis) and TrR (*x*-axis) categories, in word-initial (left) and word-medial (right) positions.

### 4.3 Discussion

The primary goal of the Perception experiment was to test how participants would categorize affricated variants of /t/. Stimuli included affricated variants of /t/ spliced from before [ɹ], concatenated with either vowel-initial (mis-matching) or [ɹ]-initial (matching) sequences. As detailed above, stimuli in these categories were treated differently by participants: affricated variants of /t/ were categorized as CH in T<sub>R</sub>V stimuli (initial: 77%, medial: 68%), but as T in T<sub>R</sub>R stimuli (initial: 81%, medial: 79%). These results suggest that, when heard outside of their natural phonetic environment, affricated variants of /t/ (in T<sub>R</sub>V stimuli) were affricated enough, and perceptually distant enough from [t], to be categorized as CH. In contrast, these same affricated variants of /t/, heard in their natural phonetic environment (in T<sub>R</sub>R stimuli), were categorized as T. In other words, listeners perceptually compensated for affrication in the environment that conditions it (e.g., Mann and Repp, 1980; Whalen, 1981; Beddor and Krakow, 1999; Smits, 2001).

As highlighted above, interpreting participants' categorizations of T<sub>R</sub>R is not straightforward because there are several possible explanations for a participant choosing TR over CHR. In an effort to mitigate some of these, the forced-choice spelling task included CH<sub>V</sub>R audio stimuli as well. It was reasoned that if participants were willing to choose CHR spellings when they heard audio stimuli with phonological [tʃ] spliced onto an [ɹ]-initial sequence, this would provide evidence that the English phonotactic bias against affricates in complex clusters was not a sufficient deterrent to prohibit CHR selections. Indeed, results showed that participants were willing to categorize CH<sub>V</sub>R audio stimuli as CH at least some of the time (initial: 45%, medial: 32%), suggesting that awareness of English phonotactics did not eliminate CHR as a viable spelling option in this task, at least for some participants.

As illustrated in Figure 4.15, participants exhibited more inter- and intra-speaker

variation in their categorizations of CH<sub>V</sub>R stimuli (particularly in word-initial position), relative to the other categories. Some speakers were systematic in their categorizations of CH<sub>V</sub>R stimuli as CH, while others systematically categorized these stimuli as T, whereas others categorized at chance. This is consistent with the idea that listeners show individual differences in perception and may differ in how much weight they assign to coarticulatory information (Shultz et al., 2012; Beddor et al., 2013). For some listeners, a [tʃ]-level amount of affrication was within their perceptual boundaries for /tɹ/ (i.e., acceptable for TR spellings).

Results were relatively consistent when comparing across word-initial and word-medial positions, with two notable exceptions. In CH<sub>V</sub>R and T<sub>R</sub>V categories, participants were less likely to categorize stimuli as CH word-medially. S24 and S32, for example, showed high CH response rates (> 75%) in word-initial position, but low CH response rates (< 25%) word-medially. Overall, this resulted in more group-level (inter-speaker) consistency for CH<sub>V</sub>R in word-medial position, with more participants categorizing tokens as T. For T<sub>R</sub>V, participant categorizations were more dispersed across the group word-medially, with more participants categorizing tokens as T. Taken together, these results suggest that affrication may be less salient in word-medial position, which may be relevant in isolating when and how the phenomenon of /tɹ/ affrication began. In a study on s-retraction with Raleigh corpus speakers, Wilbanks (2017) found that s-retraction in the speech of young females was more advanced in medial position.

# Chapter 5

## Comparing Results Across Studies

### 5.1 Comparing Production and Apparent Time

Results of the Production Study (Chapter 3) show participants producing affricate-like variants of /t/ and /d/ before /ɪ/, exhibiting coproduced gestures as early as the mid-point of the closure interval, with [tʃ]- and [dʒ]-like lip and tongue tip/blade postures, [ɪ]-like (retracted) tongue roots, and [tʃ]- and [dʒ]-like duration following release. These results provide a glimpse into how younger speakers in the Apparent Time Study (Chapter 2) may have been producing /tɪ/ and /dɪ/ sequences that resulted in higher A-Scores and A-Ratios. Results also suggest that older and younger speakers have different phonological targets that lead to acoustically different results. Although there were no older (Generation 1) participants in the Production Study, older speakers in the Apparent Time Study were producing /t/s and /d/s before /ɪ/ with higher log likelihood scores as [t]s and [d]s than [tʃ]s and [dʒ]s. Based on these findings, we might expect articulatory behaviour in older speakers that aligns with hypothesis (a) (Figure 3.35), with /t/ and /d/ before /ɪ/ more closely resembling prevocalic /t/ and /d/, both in terms of similarity between contours at the tongue tip/blade during closure, and duration of the noise interval following release. One

participant (S24) deviated from the typical group patterns at the mid-point of the closure interval, producing [t]-like tongue tip/blade postures in /tɪ/, and duration values of the noise interval following release that were mid-way between those of [t] and [tʃ]. Though it is difficult to draw conclusions based on data from one participant, these articulatory behaviours suggest different phonological targets for S24 when compared with the rest of the group, and provide a sense of how older speakers in the Apparent Time Study may have been articulating /tɪ/ and /dɪ/ sequences.

## 5.2 Comparing Perception and Production

Participants in the Production Study (Chapter 3) were a subset of the participants in the Perception Study (Chapter 4), which allowed for the comparison of production and perception results at the level of the participant. Two production measures, originally introduced in Chapter 3, were used for these comparisons. The first was Duration Ratio, which (as described in § 3.2.4) involved subtracting a participant's mean duration of the noise interval (following release) in /tɪ/ from the mean duration of the noise interval in /tʃ/, and dividing that value by the difference between the mean duration of the noise intervals in /tʃ/ and /t/. A ratio closer to 1 indicates that the participant's /tɪ/ duration is similar to their /tʃ/ duration (i.e., more [tʃ]-like) whereas a ratio closer to 0 indicates a /tɪ/ duration closer to prevocalic /t/ (i.e., more [t]-like). The second production measure was Tongue Tip/Blade Ratio, which involved comparing tongue curve differences at the tongue tip/blade. The absolute difference between each participant's /tɪ/ and /tʃ/ curves was divided by the absolute difference between their /tɪ/ and /t/ curves at the mid-point of the noise interval in /ʌ/ vowel contexts. This vowel context was chosen because in front vowel contexts, there were often no significant differences between tongue contours at the tongue tip/blade, while in back vowel contexts, where the anterior portion of

the tongue is articulatorily less encumbered, patterns between contours were visible. A ratio closer to 1 indicates that the distance between /tɹ/ and /tʃ/ is small, relative to the distance between /tɹ/ and /t/ (i.e., /tɹ/ is more [tʃ]-like than [t]-like), whereas a ratio closer to 0 indicates that the distance between /tɹ/ and /tʃ/ is similar to the distance between /tɹ/ and /t/ (i.e., /tɹ/ is as [tʃ]-like as it is [t]-like)<sup>1</sup>.

Similar perception measures were created in order to compare production and perception data. As discussed in Chapter 4, participants in the Perception Study showed low CH response rates in T<sub>V</sub>V and T<sub>V</sub>R categories (as expected), and also in T<sub>R</sub>R categories, and high CH response rates in CH<sub>V</sub>V categories (as expected), and also in TrV categories. Results were mixed in CH<sub>V</sub>R categories, with some participants exhibiting some willingness to categorize phonological affricates spliced onto [ɹ]-initial sequences as CH, others showing a stronger tendency towards T categorizations, and still others showing mixed responses. For comparing production and perception data, perception measures were created by subtracting a participant's CH response rate for affricated variants of /t/ (T<sub>R</sub>) from their CH response rate for /tʃ/ (CH<sub>V</sub>), in both vowel-initial (CH<sub>V</sub>V-T<sub>R</sub>V) and [ɹ]-initial (CH<sub>V</sub>R-T<sub>R</sub>R) contexts. These measures provided a way to gauge whether participants were categorizing affricated variants of /t/ in the same way as they were categorizing /tʃ/. A value of 0 indicates no categorization difference between the two (i.e., equivalent CH response rates for both), whereas a value of 1 indicates the two sounds were categorized differently (i.e., CH response rates in opposite directions).

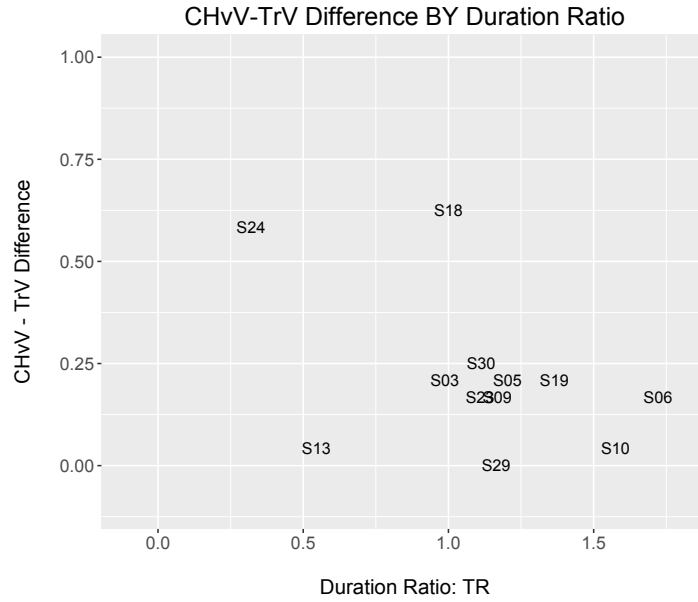
### 5.2.1 CH<sub>V</sub>V-T<sub>R</sub>V by Duration Ratio

As illustrated in Figure 5.1, which plots CH<sub>V</sub>V-T<sub>R</sub>V Differences by Duration Ratio, six participants (S03, S05, S09, S23, S29, S30) categorized T<sub>R</sub>V tokens like CH<sub>V</sub>V tokens (Difference values close to 0), and exhibited /tʃ/-like duration (Duration Ra-

---

<sup>1</sup>As stated in Chapter 3, there were no speakers who produced /tɹ/ contours closer to /t/ than /tʃ/ by the mid-point of the noise interval, but had there been, ratio values would be < 0.

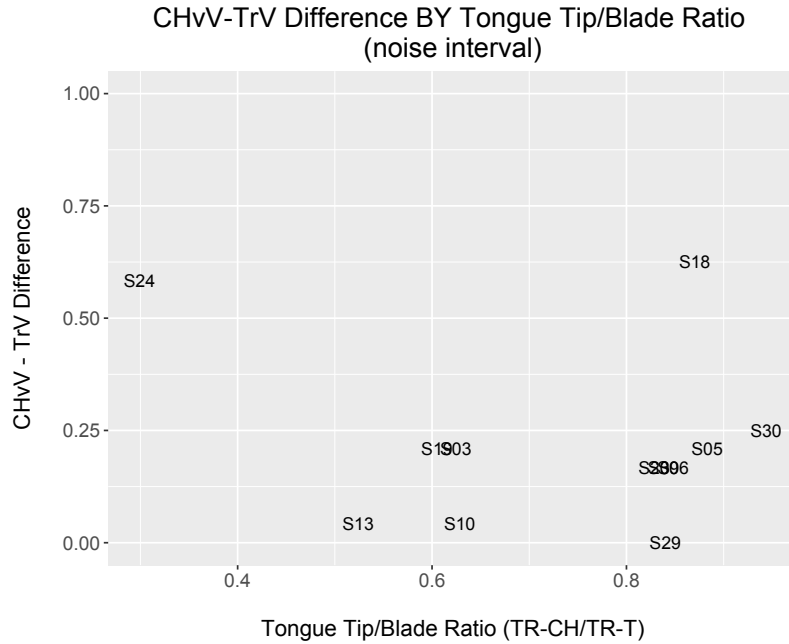
tios close 1.0). Four participants (S06, S10, S13, S19) also categorized T<sub>R</sub>V tokens like CH<sub>V</sub>V tokens, but showed more inter-speaker variation in duration, with S06, S10, and S19 exhibiting duration values greater than [tʃ]. The remaining two participants (S18 and S24) were exceptions to the group-level pattern - both categorized CH<sub>V</sub>V as CH more than they categorized T<sub>R</sub>V as CH (Difference values > 0.5), however S18 exhibited [tʃ]-like duration (ratio close to 1.0), while S24 showed more [t]-like duration (ratio < 0.5). Nevertheless, the fact that, as a group, participants generally categorized T<sub>R</sub>V tokens like CH<sub>V</sub>V tokens is, perhaps, not surprising – T<sub>R</sub>V tokens consist of affricated variants of /t/ spliced from before /ɹ/ and concatenated with vowel-initial sequences, resulting in an environment mis-match. Taken out of their natural environment, the prolonged aperiodic noise following release in T<sub>R</sub>V tokens may be more salient for all listeners, regardless of production. Given the consistency with which S18 and S24 categorized T<sub>R</sub>V tokens as T (and not CH), these participants appeared to have been attuning to something different in the aperiodic noise, that was (for them) more [t]-like. It is notable that S24 showed the least [tʃ]-like (most [t]-like) duration, and also showed less categorization similarity between /tʃ/ and affricated variants of /t/ in this context.



**Figure 5.1:** CH response rate difference for  $CH_VV$ - $T_RV$ , by Duration Ratio

### 5.2.2 $CH_VV$ - $T_RV$ by Tongue Tip/Blade

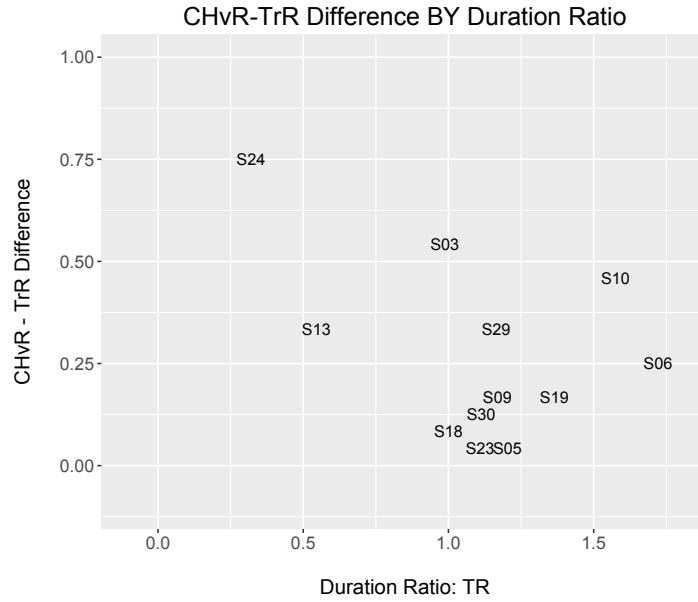
A similar pattern can be observed in Figure 5.2, which plots these same  $CH_VV$ - $T_RV$  Differences by Tongue Tip/Blade Ratio. Ten participants categorized  $T_RV$  tokens like  $CH_VV$  tokens (difference values close to 0), but six of these participants (S05, S06, S09, S23, S29, S30) showed small differences between /tɪ/ and /tʃ/ at the tongue tip/blade relative to /tɪ/ and /t/ (Tongue Tip/Blade Ratios closest to 1), while four participants (S03, S10, S13, S19) showed comparatively larger differences between /tɪ/ and /tʃ/ at the tongue tip/blade. As above, S18 and S24 were exceptions: both categorized  $CH_VV$  and  $T_RV$  differently (Difference values  $> 0.5$ ), in contrast with the rest of the group, but (consistent with his duration pattern in this context), S18 exhibited very little articulatory distance between /tɪ/ and /tʃ/ (Tongue Tip/Blade Ratio close to 1), whereas S24 showed the greatest articulatory distance between these two contours, relative to the group.



**Figure 5.2:** CH response rate difference for  $CH_VV-T_{RV}$ , by Tongue Tip/Blade Ratio

### 5.2.3 $CH_VR-T_{RR}$ by Duration Ratio

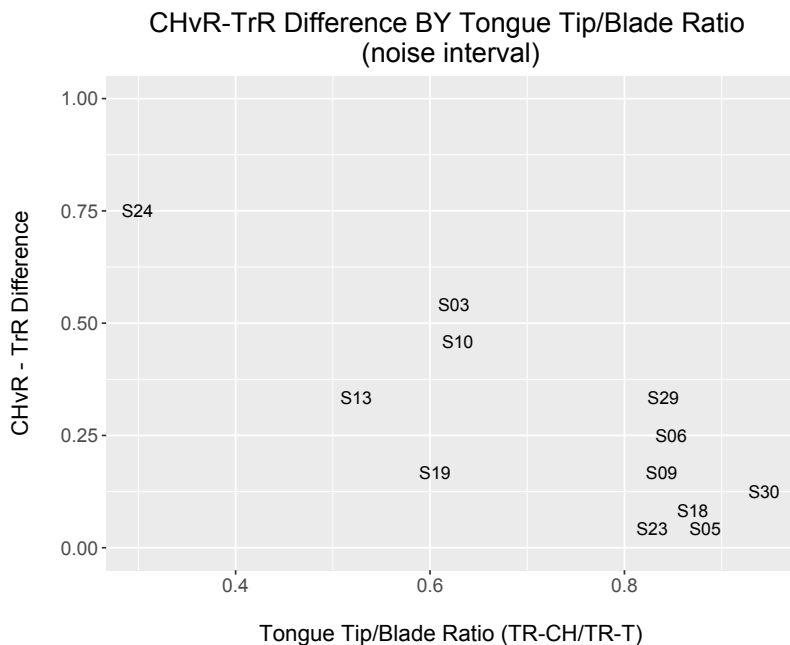
In Figure 5.3, which plots  $CH_VR-T_{RR}$  Differences by Duration Ratio, six participants (S05, S09, S18, S19, S23, S30) categorized  $T_{RR}$  tokens like  $CH_VR$  tokens (Difference values close to 0), and exhibited /tʃ/-like duration (Duration Ratios < 1.5). One participant (S06, from Delaware) also categorized  $T_{RR}$  tokens like  $CH_VR$  tokens, but showed longer duration than [tʃ] (Duration Ratio > 1.5). In contrast, one participant (S24) categorized  $T_{RR}$  and  $CH_VR$  tokens differently (generally categorizing  $T_{RR}$  tokens as T, and  $CH_VR$  tokens as CH), and also exhibited the most [t]-like duration. The remaining four participants (S03, S10, S13, S29) categorized  $T_{RR}$  and  $CH_VR$  tokens differently, but showed more inter-speaker variation in duration, though despite this variation, Duration Ratios were still more [tʃ]-like than [t]-like (with values > 0.5).



**Figure 5.3:** CH response rate difference for CH<sub>V</sub>R-T<sub>R</sub>R, by Duration Ratio

#### 5.2.4 CH<sub>V</sub>R-T<sub>R</sub>R by Tongue Tip/Blade Ratio

In Figure 5.4, which plots CH<sub>V</sub>R-T<sub>R</sub>R Differences by Tongue Tip/Blade Ratio, six participants (S05, S06, S09, S18, S23, S30) categorized T<sub>R</sub>R tokens like CH<sub>V</sub>R tokens (Difference values close to 0), and exhibited very little articulatory distance between /tɪ/ and /tʃ/ relative to /tɪ/ and /t/ at the tongue tip/blade (Ratio values close to 1). One participant (S19) also categorized T<sub>R</sub>R and CH<sub>V</sub>R tokens similarly, but exhibited greater articulatory distance between /tɪ/ and /tʃ/, while another participant (S29) showed less categorization similarity between T<sub>R</sub>R and CH<sub>V</sub>R tokens, but exhibited minimal articulatory distance between /tɪ/ and /tʃ/. Three participants (S03, S10, S13) showed less categorization similarity between T<sub>R</sub>R and CH<sub>V</sub>R tokens, and also exhibited comparatively greater articulatory distance between /tɪ/ and /tʃ/. The remaining participant (S24) showed the greatest differences between /tɪ/ and /tʃ/ relative to /tɪ/ and /t/ at the tongue tip/blade, and also showed the least amount of categorization similarity between CH<sub>V</sub>R and T<sub>R</sub>R.



**Figure 5.4:** CH response rate difference for  $CH_VR-T_{RR}$ , by Tongue Tip/Blade Ratio

### 5.2.5 Discussion

The production-perception comparisons presented in Figures 5.1 through 5.4 illustrate a number of different patterns. As a group, participants typically categorized  $T_{RV}$  tokens like  $CH_VV$  tokens, regardless of production. This was not completely unexpected – when the environment that conditions the affrication is removed (as it is with  $T_{RV}$  tokens), listeners have no way to perceptually compensate for the affrication, which may make it more salient.

As a group, participants also categorized  $T_{RR}$  and  $CH_VR$  tokens similarly, but exhibited more inter-speaker variation in this context. Some participants showed very little difference in their responses to  $T_{RR}$  and  $CH_VR$  tokens (generally categorizing both tokens as T) and typically demonstrated more [tʃ]-like production, while others categorized  $T_{RR}$  and  $CH_VR$  tokens differently, and often showed correspondingly less articulatory similarity between /tɪ/ and /tʃ/.

Taken together, comparisons of perception and production data demonstrate

that while participants generally categorized [tʃ] (CH<sub>V</sub>) and affricated variants of /t/ (T<sub>R</sub>) similarly, a relationship between perception and production was only apparent when variants were concatenated with [ɹ]-initial sequences, where the trigger for /tɹ/ affrication was also present. In other words, taken out of their natural environment, affricated variants of /t/ were categorized as [tʃ], but placed within the context that conditions the affrication, there was an apparent correlation between how these variants were categorized and how /tɹ/ sequences were produced. As illustrated in Table 5.1, (with some exceptions, e.g., S29) participants producing more [tʃ]-like /tɹ/ categorized affricated variants of /t/ and /tʃ/ more similarly (Subgroup 1), while participants producing less [tʃ]-like /tɹ/ categorized these sounds as more different from one another (Subgroup 2)<sup>2</sup>. These results suggest that differences in a participant's perceptual boundaries for /tɹ/ may shape and be shaped by their production (e.g., Liberman and Mattingly, 1985; Hay et al., 2006; also, Galantucci et al., 2006 for a review of the motor theory of speech perception). This idea is consistent with electromyography research conducted by Bell-Berti et al. (1979), who found that speakers employed one of two articulatory strategies in distinguishing between /i ɪ/ and /e ε/ vowel pairs (exploiting either height or tenseness differences), and that these same speakers, as listeners, grouped along the same lines during a perception task with /i ɪ/ vowels.

The inter-speaker comparisons presented here point to differences in perceptual boundaries for each subgroup: Subgroup 1 participants were willing to accept [tʃ] as a possible candidate for /t/ in a pre-[ɹ] environment, whereas Subgroup 2 participants were not nearly as willing. Further, for the Subgroup 2 participants, who categorized T<sub>R</sub>R and CH<sub>V</sub>R tokens differently (Table 5.1), there was also inter-speaker variation in terms of *how* these tokens were categorized by participants as

---

<sup>2</sup>These participant subgroupings were also apparent in Figure 3.34, which compared both production measures (Tongue Tip/Blade Ratio by Duration Ratio) in a single plot. S29 was the exception (as can be observed in Figure 3.34), patterning with Subgroup 1 participants for both production measures, but deviating (slightly) from Subgroup 1 in perception.

different.

|                   | <i>Participant</i> | <b>Production</b>          |                          | <b>Perception</b>                              |
|-------------------|--------------------|----------------------------|--------------------------|--|
|                   |                    | <i>Duration:</i> [tʃ]-like | <i>Tongue:</i> [tʃ]-like | <i>CH<sub>V</sub>R-T<sub>R</sub>R:</i> similar |
| <b>Subgroup 1</b> | S05                | ✓                          | ✓                        | ✓  |
|                   | S06                |                            | ✓                        | ✓  |
|                   | S09                | ✓                          | ✓                        | ✓  |
|                   | S18                | ✓                          | ✓                        | ✓  |
|                   | S19                | ✓                          |                          | ✓  |
|                   | S23                | ✓                          | ✓                        | ✓  |
|                   | S30                | ✓                          | ✓                        | ✓  |
| <b>Subgroup 2</b> | S03                | ✓                          |                          |  |
|                   | S10                |                            |                          |  |
|                   | S13                |                            |                          |  |
|                   | S24                |                            |                          |  |
|                   | S29                | ✓                          | ✓                        |  |

**Table 5.1:** Production-perception comparisons. A participant’s Duration Ratio is categorized as [tʃ]-like if it falls between 1.0 and 1.5 and Tongue Tip/Blade Ratio is categorized as [tʃ]-like if it falls between 0 and 0.2. A participant’s perception of T<sub>R</sub>R and CH<sub>V</sub>R tokens is categorized as similar, if their Difference value falls at or below 0.25.

S24, for example, categorized T<sub>R</sub>R tokens as T 100% of the time and CH<sub>V</sub>R tokens as CH 75% of the time. S29 categorized T<sub>R</sub>R tokens as T 75% of the time, but CH<sub>V</sub>R tokens as CH only 55% of the time (and as T 45% of the time), suggesting greater overlap in S29’s perceptual boundaries for affricated variants of /t/ and /tʃ/ (as compared with S24)<sup>3</sup>. In fact, S24 was one of the only participants to consistently categorize CH<sub>V</sub>R tokens as CH (S10 was the other), and she was also one of the only participants to categorize T<sub>R</sub>V tokens as T (S18 was the other). These interspeaker differences demonstrate that S24 was able to distinguish between T<sub>R</sub> and CH<sub>V</sub> tokens (in both contexts), while other participants were not. In her language

<sup>3</sup>See Figure 4.18 for participants’ raw CH response rates in this context.

background questionnaire (Appendix B.2), S24 reported “a slight lisp sometimes” (§ 3.1.1) and, although no lisp was detectable in her recorded speech, this may have informed how she approached the production task. S24’s speech was not noticeably different from the other participants – she spoke at a rapid pace, without any obvious hyperarticulation, but if (as a result of her background) she possessed (for example) more awareness of articulatory movements during production, this may have influenced her approach to the production and perception tasks. Nevertheless, across measures, when compared with the other participants, S24 showed the least amount of articulatory similarity between */tɪ/* and */tʃ/*, and the least amount of categorization similarity between */tɪ/* and */tʃ/* as well, reinforcing the notion of a production-perception link, but suggesting a different overall phonological orientation from the rest of the group.

# Chapter 6

## General Discussion

This dissertation was designed to investigate a reported sound pattern in North American English: the affrication of /t/ and /d/ before /ɪ/. The goal of the research was to explore whether coronal stops that are coarticulated with /ɪ/ in /tɪ/ and /dɪ/ sequences simply *sound* like affricates, or whether speakers are producing distinct articulatory targets and have phonologized a coarticulatory effect. The exploration was conducted over three separate studies, aimed at answering three specific questions:

**Q1: *Apparent Time Study*** – Is the phenomenon of /tɪ/ and /dɪ/ affrication in English a sound change in progress?

**Q2: *Production Study*** – Are English /t/ and /d/ in /tɪ/ and /dɪ/ clusters articulated like prevocalic [t] and [d], like prevocalic [tʃ] and [dʒ], like neither, or like both?

**Q3: *Perception Study*** – Do English speakers categorize affricated variants of /t/ found in /tɪ/ clusters as T or CH?

In the Apparent Time Study (Chapter 2), younger (Generation 3) speakers born after the middle of the 20th century, produced /tɪ/ and /dɪ/ sequences that were

more like [tʃ] and [dʒ], while older (Generation 1) speakers born early in the 20th century, produced /tɹ/ and /dɹ/ sequences that were more like [t] and [d]. Females were leading the change, which was underway near the middle of the 20th century. Further, the change in progress appears to be phonetically gradual, and not the abrupt introduction of a novel variant as a result of dialect contact.

Production Study results (Chapter 3) were in keeping with these findings: participants (who were in the same age range as Generation 3 speakers from the apparent time study) coproduced /tɹ/ and /dɹ/ sequences that were at once part post-alveolar affricate and part [ɹ]. Articulatory data showed participants exhibiting affricate-like lip protrusion at the start of the closure interval in /tɹ/ and /dɹ/ sequences, and not the lip rounding we would expect to see if the early lip engagement were due to coarticulation with [ɹ]. In other words, participants in the Production study had phonologized what was once a coarticulatory effect (the aftermath of coarticulation). Affricate-like lip protrusion was also observed at the mid-point of the noise interval and at the start of the /ɹ/ gesture in /tɹ/ and /dɹ/ sequences. Participants showed retracted tongue roots at the mid-point of the closure interval, which points to early articulation of the /ɹ/ gesture. This is in keeping with the notion that early articulation of a complex gesture like /ɹ/ is possible on an articulatory tier where there is no articulatory competition (i.e., neither /t/ nor /tʃ/ engage the tongue root). At the tongue tip/blade, /tɹ/ and /dɹ/ postures were affricate-like and significantly different from prevocalic /t/ and /d/ at the mid-point of the closure interval, and at the mid-point of the noise interval where participants also exhibited ([ɹ]-like) concavity at the tongue dorsum. [tʃ]-like and [dʒ]-like tongue tip/blade postures were consistent with the [tʃ]-like and [dʒ]-like lip postures observed at these same time points – participants were producing affricated variants of /t/ and /d/ before /ɹ/ that more closely resembled post-alveolar affricates. Duration of the noise interval in /t/ and /d/ before /ɹ/ was also similar to prevocalic /tʃ/ and /dʒ/, and significantly

different from prevocalic /t/ and /d/.

In the Perception Study (Chapter 4), results suggest that affricated variants of /t/, when heard outside of their natural phonetic environment (i.e., concatenated with vowels) were affricated enough, and perceptually distant enough from [t], to be categorized as CH. In contrast, these same affricated variants of /t/, heard in their natural phonetic environment were categorized as T. In other words, listeners perceptually compensated for affrication in the environment that conditions it (e.g., Mann and Repp, 1980; Whalen, 1981; Beddor and Krakow, 1999; Smits, 2001). Participants also exhibited a willingness to accept phonological affricates concatenated with [ɹ] as candidates for TR spellings.

While comparing results from across studies (§ 5.2), a pattern emerged between participants' production and perception of /tɹ/ sequences, prompting the division of participants into two subgroups (Table 5.1). Subgroup 1 participants perceived affricated variants of /t/ spliced from before /ɹ/ and phonological affricates similarly in pre-[ɹ] environments, and produced /tɹ/ sequences as post-alveolar affricates, exhibiting little acoustic and articulatory distance from phonological /tʃ/. Subgroup 2 participants perceived affricated variants of /t/ spliced from before /ɹ/ and phonological affricates less similarly in pre-[ɹ] environments, and showed more variation in production, producing post-alveolar affricates, but exhibiting greater acoustic and articulatory differences from phonological /tʃ/. These results suggest that differences in a participant's perceptual boundaries for /tɹ/ may be shaped or be shaped by their production (e.g., Liberman and Mattingly, 1985; Bell-Berti et al., 1979; Galantucci et al., 2006). Despite these differences (and with one exception), participants in both subgroups produced /tɹ/ sequences that were more [tʃ]-like than [t]-like, with production targets that were distinct from /t/, suggesting both subgroups have phonologized a coarticulatory effect, though in different ways. For Subgroup 1 participants, there may be no phonological difference between affricated

variants of /t/ and /d/ before /ɹ/ and phonological affricates, whereas for Subgroup 2 participants (except S24), there may be new target entries, distinct from both phonological affricates and coronal stops. We might imagine two sets of rules:

**Subgroup 1 Rule:** t/d → tʃ/dʒ / \_ ɹ

**Subgroup 2 Rule:** t/d → post-alveolar affricate (distinct from tʃ/dʒ) / \_ ɹ

For Subgroup 1 participants, we could characterize this as a phonological merger, though not in the traditional sense, since no words are being merged and no contrasts are being neutralized. Instead, the merger involves sequences that did not already occur: /tʃɹ/ and /dʒɹ/. There is no expectation that [tʃ] and [dʒ] before [ɹ] will exhibit behaviour that is phonetically identical to [tʃ] and [dʒ] in other contexts - allophones are not phonetically identical. In this sense, the fact that we see articulatory differences between affricated variants of /t/ and /d/ before /ɹ/ and prevocalic [tʃ] and [dʒ], as well as differences in how these types of sounds are categorized when spliced out of their natural phonetic environments is not inconsistent with a characterization of phonological merger.

In their study on American English s-retraction, Baker et al. (2011:360) found that while only 8 of their 19 participants were classified as “retractors” (producing e.g., [ʃtrit] rather than [strit]), all participants (including “non-retractors”) exhibited “a substantial coarticulatory bias toward s-retraction”. Represented as a ratio, retractors were producing [s] with a centroid frequency that was 75.5% of the way to canonical [ʃ], but non-retractors were producing [s] with a centroid frequency that was 65% of the way toward the retracted [s] of the retractors (or 47% of the way to [ʃ]). Using a similar approach based on Tongue Tip/Blade Ratio, which (as discussed in § 3.2.4 and § 5.2) calculates articulatory distance between /tɹ/ and /tʃ/ relative to /tɹ/ and /t/ at the tongue tip/blade, participants in Subgroup 2 (with

the exception of S24<sup>1</sup>) exhibited tongue postures that were 59.5% of the way from [t] to [tʃ], as compared with Subgroup 1, whose tongue postures were 86.5% of the way to [tʃ]. Expressed as a ratio ( $59.5\%/86.5\%=68.9\%$ ), participants in Subgroup 2 were producing /tɹ/ sequences that were 69% of the way towards those of Subgroup 1 (in terms of similarity to [tʃ] at the tongue tip/blade). Although S24 patterned with Subgroup 2, her production and perception behaviours were far less [tʃ]-like than the rest of her subgroup. Calculated separately, S24 exhibited tongue postures that were 29.9% of the way to [tʃ] at the tongue tip/blade, or 34.5% of the way toward the [tʃ]-like /tɹ/ sequences of participants in Subgroup 1, which is consistent with the notion that S24, has maintained /t/ as a phonological target (and has *not* phonologized a coarticulatory effect).

Many approaches to sound change have in common the notion that change is due to speaker- and/or listener-based errors (e.g., Paul, 1880; Ohala, 1981, 1983; Pierrehumbert, 2002). Baker et al. (2011) have argued that accumulation-of-error theories are problematic because they over-predict sound change and do not provide an adequate account for when sound change fails to occur. This is the crux of the actuation problem, as framed by Weinreich et al. (1968). In keeping with Janda and Joseph (2003), who argue that the initiation of a sound change is a brief and phonetically sudden event (a ‘burst’), Baker et al. (2011:351) propose that “the potential for sound change arises when inter-speaker variability is so great that some speakers actually produce a sound that can be perceived as a distinct target”. Further, the authors argue that the actuation of a sound change is rare because it depends on “the chance alignment of extreme coarticulation with extreme influence”. In other words, exaggeration of a (phonetically-motivated) coarticulatory effect can be perceived by a listener as a novel target. If this new target is adopted by the

---

<sup>1</sup>In many respects, S24 could form her own subgroup, but was included with Subgroup 2 because, despite more extreme behaviours (greater articulatory distance from [tʃ], more /t/-like duration, etc.), her overall production and perception patterns were consistent with the rest of her subgroup.

listener-turned-speaker, phonologization has occurred. If this speaker is influential within his/her speech community, a sound change can begin to spread (see also Janda, 1999; Janda and Joseph, 2003).

What factors might have motivated a change in target for English /tɹ/ and /dɹ/ sequences? Ohala and Solé (2010:47) refer to “emergent affricates” that result from the extended period of turbulence as a stop is released into a segment with a high tongue position. In their framing, increased frication noise occurs during the early portion of a segment with narrow constriction, but can be interpreted by a listener as belonging to the preceding stop (stop + prolonged frication noise = affricate), rather than to the following, narrowly constricted segment. While participants’ longer noise interval duration before /ɹ/ (relative to prevocalic /t/) could not be accounted for by aerodynamic factors alone (§ 3.3), at an earlier stage of change (or in the speech of older speakers), prolonged frication noise may have led some listeners to perceive phonetic [t]s coarticulated with [ɹ] as affricated.

Perception Study results showed that participants were willing to categorize phonological /tʃ/ concatenated with [ɹ] as CHR at least some of the time (initial: 45%, medial: 32%), suggesting that awareness of English phonotactics did not eliminate CHR as a viable spelling option in the task, at least for some participants.

Children are also able to perceive affrication on variants of /t/ in contexts before [ɹ], but in contrast with the adults in this dissertation research, systematically represent words like *tree* with <chr> and *dragon* with <jr> before they have learned the standardized English spellings (Read and Treiman, 2012). If children interpret these sounds as phonological affricates during the period of language acquisition, it is plausible that, for some children, this phonological representation may influence production and persist into adulthood. Depending on social influence, these now adult production strategies could provide the seeds for change. In addition to the evidence of children’s perceptual awareness of /tɹ/ and /dɹ/ affrication, there

is anecdotal evidence to suggest that (at least some) adults are also aware of the affrication in /tɹ/ and /dɹ/ sequences and use English spelling to reflect what they hear. For example, a Google search for “*chru dat*” (a variation of “*true that*”) resulted in 46,700 results in 0.36 seconds, and while some hits were for a person named *C.H. Rudat*, most of the top hits appeared to be marking an awareness of English sounds. Further, there is evidence of this awareness at a juicery in Maryland called “*JRINK*”, a clothing shop on Twitter called “*the jrive*” that has been “*jriven to succeed since ‘96*”, and in the spelling of NBA point guard *Jrue Holiday*’s name. You can find YouTube videos entitled e.g., “*I’m a big monster chruck fan*” and “*Chrash Chruck*”, Urban Dictionary entries for e.g., “*chru*”, “*chruck*” and “*jrunck*”, and seemingly endless discussions on *wordreference.com* and *reddit.com* about the pronunciation of these and other /tɹ/ and /dɹ/ words.

As observed in the Perception Study, some participants categorized audio tokens of phonological /tʃ/ and affricated variants of /t/ concatenated with [ɹ] as TR. In other words, for some listeners, a [tʃ]-level amount of affrication was within their perceptual boundaries for /tɹ/. Affricates in complex onsets are phonotactically restricted in English, such that [t] and [tʃ] are not contrastive before /ɹ/. This means there is no semantic consequence of producing an affricated variant in this context: [t.ɹ] and [tʃ.ɹ] may both be parsed as *tree* (whereas [ti] and [tʃi] are meaningfully different). Listener-tolerated inter-speaker variation during production of /tɹ/ and /dɹ/ may provide the ideal conditions for change. If some speakers produce more affricated variants that are tolerated by some listeners and, as Baker et al. (2011) have proposed, perceived by others as a distinct target, the potential for change (through adoption of a new target) exists.

Another factor that may have motivated a change in target is what Fowler and Saltzman (1993:182) refer to as “blending strength”. As described by Farnetani and Recasens (2010), “gestures with a high degree of blending strength resist interfer-

ence from other gestures and at the same time themselves induce strong coarticulatory effects”. As discussed in Chapter 1, although Fowler and Saltzman (1993) do not explicitly reference English /ɹ/, their blending strength ranking appears to assign greater influence to consonants “requiring extreme constrictions and/or placing strong constraints on articulator movements”. Similarly, Recasens et al.’s (1997) DAC model posits that lingual coarticulation is determined by the demands placed on the tongue during production – vowels and consonants requiring the greatest degree of lingual precision receive the highest values, where a high DAC value means greater coarticulatory resistance and greater coarticulatory influence over adjacent segments. In a study comparing the blending strength of English liquids, Walker, Proctor, Smith, and Enzinger (2016) argued that /ɹ/ is “defined by a tongue body gesture with stronger blending parameters than that of the lateral, which results in its exerting a more global influence on tongue shaping and constriction location than /l/”. The authors found that, despite variability in /ɹ/ production, the various tongue postures for /ɹ/ acted to “constrain global tongue shaping more consistently across different positions in the syllable” when compared with /l/. In the Production Study, [ɹ]’s influence was apparent: tongue root retraction for the /ɹ/ was evident as early as the mid-point of the closure interval for /t/ and /d/, and concavity at the tongue dorsum (a characteristic of a bunched English /ɹ/) as early as the mid-point of the noise interval following release. Further, participants tongue tip/blade postures were less like coronal stops and more like /tʃ/ and /dʒ/, which are articulatorily more similar to a bunched /ɹ/ (Gick, 1999)<sup>2</sup>.

Löfqvist (1990:316) describes how the blending of gestures may be advantageous for reasons of “computational simplicity”. This idea has been echoed in work by Guenther (1994), who proposed that speakers actively optimize gestural overlap for reasons of articulatory efficiency, similar to Lindblom’s (1983) “economy of effort”.

---

<sup>2</sup>Gick (1999) refers to articulatory similarity between bunched [ɹ] and [ʃ].

In /tɹ/ and /dɹ/ sequences, early retraction of the tongue root in anticipation of the upcoming /ɹ/ gesture is consistent with this idea of efficiency. Rosenbaum and Jorgensen (1992) refer to what they call *end-state-comfort* – a concept that extends beyond linguistics, and focuses on gestures more broadly. The general idea is that people think about how to initiate a movement with the end of the movement in mind. In the context of /tɹ/ and /dɹ/, [tʃ] and [dʒ] are articulatorily more similar to [ɹ] (Gick, 1999) than [t] is to [ɹ]. Further, [ɹ] is thought to have more blending strength than [t] or [d], exerting more influence over adjacent segments. Initiating an oral constriction for /t/ or /d/ may achieve more end-state-comfort if the constriction is produced farther back than the alveolar ridge, at a place of articulation that is nearer to, or consistent with, the constriction location for the upcoming and comparatively complex post-alveolar approximant.

Although, to the best of my knowledge, there has been no experimental work investigating the affrication of English /tɹ/ and /dɹ/, there is a large body of work investigating the production of English /ɹ/ (e.g., Delattre and Freeman, 1968; Lindau, 1985; Westbury et al., 1998; Guenther et al., 1999; Campbell et al., 2010; Baker et al., 2011; Magloughlin, 2016). Campbell et al. (2010) have explored the articulatory timing of the gestures for /ɹ/, an articulatorily complex speech sound with three points of constriction: at the lips, at the palate (tongue blade), and at the pharyngeal wall (tongue root). As reported by Campbell et al. (2010:66, Table 4), previous research on gestural timing of /ɹ/ in initial position has been inconsistent, with some studies reporting simultaneous production of all three gestures (Sproat and Fujimura, 1993; Browman and Goldstein, 1995; Gick et al., 2006), and others reporting the lips and tongue blade patterning together (Gick and Campbell, 2003), or the lips preceding the tongue blade preceding the tongue root (Campbell et al., 2010). Production Study participants producing /tɹ/ and /dɹ/ sequences showed evidence of tongue root engagement as early as the mid-point of the closure inter-

val for /t/ and /d/, and preceding tongue blade (and tongue dorsum) engagement, with little or no /ɹ/-like lip engagement, presumably due to conflict with the preceding gesture for control of the lips articulator. These findings, though not directly comparable to previous studies looking at initial and final /ɹ/, add to the body of research on articulator timing for this complex gesture.

## 6.1 Conclusion

In this dissertation, I have argued that the affrication of /t/ and /d/ before /ɹ/ in North American English is an active sound change in progress that has been phonologized. The investigation was conducted in Raleigh, North Carolina over three separate studies. Apparent Time Study results showed a sound change underway for English speakers by the middle of the 20th century, with females leading the change. Speakers born early in the 20th century produced /tɹ/ and /dɹ/ sequences that were more like [t] and [d], while speakers born after the middle of the 20th century produced /tɹ/ and /dɹ/ sequences that were more like [tʃ] and [dʒ]. Crucially, the change in progress showed a gradual phonetic change, and not the abrupt introduction of new variants imported into the community through dialect contact. Consistent with these findings, Production Study results showed participants were not just coarticulating [tɹ] and [dɹ] sequences, but rather, had phonologized a coarticulatory effect, producing targets that were [tʃ]-like and [dʒ]-like (and distinct from prevocalic [t] and [d]), coproduced with [ɹ]: the aftermath of coarticulation. Perception Study results showed participants categorized affricated variants of /t/ as T in pre-[ɹ] contexts, exhibiting perceptual compensation in the environment that conditions the affrication, but categorized these same affricated variants of /t/ as CH before vowels, when taken outside of their natural phonetic environment. Participants also categorized affricated variants of /t/ and phonological /tʃ/ similarly,

with many exhibiting a willingness to accept phonological affricates as candidates for TR spellings. Taken together, these findings indicate that the phenomenon of /tɹ/ and /dɹ/ affrication has been phonologized.

# Appendix A

## Experimental Stimuli

### A.1 Production Stimuli and Filler

Table A.1: Production Stimuli

|          | /tɪ/  | /t/   | /tʃ/  | /dɪ/  | /d/  | /dʒ/  |
|----------|-------|-------|-------|-------|------|-------|
| <b>i</b> | tree  | tea   | cheap | dream | deem | jeans |
|          | treat | teach | cheek | Dreep | deep | jeep  |
| <b>ɪ</b> | trick | tick  | chick | drip  | dip  | Jim   |
|          | trip  | tip   | chip  | drift | dibs | jib   |
| <b>ʌ</b> | truck | tuck  | chuck | drug  | duck | jug   |
|          | trust | tub   | chum  | drum  | dumb | jump  |
| <b>ɑ</b> | trot  | talk  | chalk | drop  | dock | job   |
|          | Tron  | top   | chop  | draw  | dog  | jock  |

**Table A.2:** Production Filler

|                |                 |               |                  |
|----------------|-----------------|---------------|------------------|
| administration | bone            | kiss rock     | shout            |
| astronomy      | bony            | leap          | shut             |
| Australia      | bore sip        | left          | sip              |
| babe           | bore zip        | lip           | sob              |
| back           | bores           | loop          | sound            |
| bad            | both            | morse         | south            |
| bade           | bud             | oath          | steep            |
| bag            | bun             | only          | stick            |
| bail           | call a puppy    | parse         | stop             |
| bam            | call the puppy  | peel          | strawberry       |
| ban            | destruction     | peep          | streaky          |
| bane           | distract        | pierce        | street           |
| bang           | distributed     | pill          | strictly         |
| bar sip        | endless trust   | pip           | strip            |
| bar zip        | eucalyptus tree | pole          | strong           |
| bars           | fabulous Tron   | pool          | struck           |
| bathrobe       | fizz            | poop          | struggle         |
| bathroom       | fizz reed       | pull          | stuck            |
| bead           | fizz road       | reed          | sub              |
| bean           | fizz rock       | ref           | sumptuous treat  |
| beb            | furious trot    | registration  | these            |
| bed            | heap            | rip           | thirty           |
| beer sip       | heft            | road          | thorough         |
| beer zip       | hip             | rock          | though           |
| beers          | home            | rub           | three            |
| bell           | hoop            | rude          | threw            |
| Ben            | hope            | saw a puppy   | throw            |
| bid            | illustrated     | saw the puppy | vegas trip       |
| bin            | ingenious trick | seep          | windowless truck |
| bod            | kiss            | sheep         | zip              |
| bode           | kiss reed       | ship          |                  |
| bon            | kiss road       | shop          |                  |

## A.2 Perception Non-Word Stimuli and Filler

Table A.3: Perception Stimuli

| Vowel Context | Category                 |                          |
|---------------|--------------------------|--------------------------|
|               | T <sub>V</sub> V tokens  | T <sub>V</sub> R tokens  |
| [ɪ]           | T <sub>V</sub> IV        | T <sub>V</sub> RIV       |
|               | T <sub>V</sub> IB        | T <sub>V</sub> RIB       |
|               | MET <sub>V</sub> IFFEN   | MET <sub>V</sub> RIFFEN  |
|               | VET <sub>V</sub> IBBEN   | VET <sub>V</sub> RIBBEN  |
| [ʌ]           | T <sub>V</sub> UV        | T <sub>V</sub> RUV       |
|               | T <sub>V</sub> UD        | T <sub>V</sub> RUD       |
|               | MET <sub>V</sub> UPPEN   | MET <sub>V</sub> RUPPEN  |
|               | VET <sub>V</sub> UZZEN   | VET <sub>V</sub> RUZZEN  |
| [ɑ]           | T <sub>V</sub> OV        | T <sub>V</sub> ROV       |
|               | T <sub>V</sub> OB        | T <sub>V</sub> ROB       |
|               | MET <sub>V</sub> OFFEN   | MET <sub>V</sub> ROFFEN  |
|               | VET <sub>V</sub> OBHEN   | VET <sub>V</sub> ROBBEN  |
| Vowel Context | T <sub>R</sub> V tokens  | T <sub>R</sub> R tokens  |
| [ɪ]           | T <sub>R</sub> IV        | T <sub>R</sub> RIV       |
|               | T <sub>R</sub> IB        | T <sub>R</sub> RIB       |
|               | MET <sub>R</sub> IFFEN   | MET <sub>R</sub> RIFFEN  |
|               | VET <sub>R</sub> IBBEN   | VET <sub>R</sub> RIBBEN  |
| [ʌ]           | T <sub>R</sub> UV        | T <sub>R</sub> RUV       |
|               | T <sub>R</sub> UD        | T <sub>R</sub> RUD       |
|               | MET <sub>R</sub> UPPEN   | MET <sub>R</sub> RUPPEN  |
|               | VET <sub>R</sub> UZZEN   | VET <sub>R</sub> RUZZEN  |
| [ɑ]           | T <sub>R</sub> OV        | T <sub>R</sub> ROV       |
|               | T <sub>R</sub> OB        | T <sub>R</sub> ROB       |
|               | MET <sub>R</sub> OFFEN   | MET <sub>R</sub> ROFFEN  |
|               | VET <sub>R</sub> OBHEN   | VET <sub>R</sub> ROBBEN  |
| Vowel Context | CH <sub>V</sub> V tokens | CH <sub>V</sub> R tokens |
| [ɪ]           | CH <sub>V</sub> IV       | CH <sub>V</sub> RIV      |
|               | CH <sub>V</sub> IB       | CH <sub>V</sub> RIB      |
|               | MECH <sub>V</sub> IFFEN  | MECH <sub>V</sub> RIFFEN |
|               | VECH <sub>V</sub> IBBEN  | VECH <sub>V</sub> RIBBEN |
| [ʌ]           | CH <sub>V</sub> UV       | CH <sub>V</sub> RUV      |
|               | CH <sub>V</sub> UD       | CH <sub>V</sub> RUD      |
|               | MECH <sub>V</sub> UPPEN  | MECH <sub>V</sub> RUPPEN |
|               | VECH <sub>V</sub> UZZEN  | VECH <sub>V</sub> RUZZEN |
| [ɑ]           | CH <sub>V</sub> OV       | CH <sub>V</sub> ROV      |
|               | CH <sub>V</sub> OB       | CH <sub>V</sub> ROB      |
|               | MECH <sub>V</sub> OFFEN  | MECH <sub>V</sub> ROFFEN |
|               | VECH <sub>V</sub> OBHEN  | VECH <sub>V</sub> ROBBEN |

**Table A.4:** Perception Filler

| <b>Vowel Context</b> | <b>Filler Spelling Pairs</b>                                 |
|----------------------|--|
| [ɪ]                  | DLIV ~ TLIV<br>DLIF ~ TLIF                                   |
| [ʌ]                  | GNUB ~ KNUB<br>GNUP ~ KNUP<br>VNUD ~ FNUD<br>VNUT ~ FNUT     |
| [ɑ]                  | BNOG ~ PNOG<br>BNOCK ~ PNOCK                                 |
| [u]                  | VOOB ~ FOOB<br>VOOP ~ FOOP<br>GLOOB ~ KLOOB<br>GLOOP ~ KLOOP |

# Appendix B

## Consent Form and Questionnaire

### B.1 University of Ottawa Consent Form



Université d'Ottawa  
Faculté des arts  
Linguistique  
University of Ottawa  
Faculty of Arts  
Linguistics

☎ 613-562-5286  
☎ 613-562-5141  
70 Laurier E.  
Ottawa ON K1N 6N5 Canada  
www.uOttawa.ca

## INFORMED CONSENT FORM for RESEARCH

### Sounds of Speech – Doctoral Research Project

**Doctoral Researcher:** Lyra Magloughlin  
PhD Candidate, University of Ottawa

**Doctoral Supervisor:** Dr. Jeff Mielke  
Adjunct Professor, University of Ottawa  
Associate Professor, North Carolina State University

#### **What are some general things you should know about research studies?**

You are being asked to take part in a research study. Your participation in this study is voluntary. You have the right to be a part of this study, to choose not to participate or to stop participating at any time without penalty. The purpose of research studies is to gain a better understanding of a certain topic or issue. You are not guaranteed any personal benefits from being in a study. Research studies also may pose risks to those that participate. In this consent form you will find specific details about the research in which you are being asked to participate. If you do not understand something in this form it is your right to ask the researcher for clarification or more information. A copy of this consent form will be provided to you. If at any time you have questions about your participation, do not hesitate to contact the researcher(s) named above.

#### **What is the purpose of this study?**

This research aims to investigate how people produce speech, and how sound is used in language, using audio recording and different techniques for measuring the movements of speech organs such as the tongue and lips.

#### **What will happen if you take part in the study?**

If you agree to participate in this study, you will be asked to spend up to  60 /  75 minutes in the laboratory (located at \_\_\_\_\_) and do some or all of the following activities (all the ones that are checked):

- Read aloud from a word list, and/or a list of sentences, and/or repeat after speech you hear on headphones or speakers, while being audio recorded using a head-worn or stand-mounted microphone.
- Listen to a series of unfamiliar words over headphones, and make selections on a computer based on what you hear.
- Be video recorded while speaking.
- Have your tongue movements recorded by an ultrasound machine while you speak. The ultrasound probe covered with gel will be placed under your chin and held in place using a specially-designed headset or a stand. You may be asked to wear glasses, goggles, a headband, or a hair clip with a small piece of wood or accelerometers attached, and/or to lean against a head-rest, instead of wearing a headset. You may be asked to visually align two objects to help maintain a consistent head position. You may be asked to drink water and hold a tongue depressor between your teeth (not at the same time) as part of this process, in order to allow the ultrasound machine to create an image of the top of your mouth and your bite plane.
- Fill out a questionnaire about your language background and related information.

**Risks**

You may feel physically uncomfortable due to sitting in the same position for a long time or due to wearing equipment that you do not wear in everyday life. Great care will be taken to assure that the equipment is positioned so as to maximize your comfort before proceeding with the experiment. You can alert the experimenter if you feel any discomfort.

Any time you touch an object that has been touched by another person, there is a risk of being exposed to viruses or bacteria. Laboratory instruments that come into contact with people (transducer, head-rest, etc.) are sanitized between uses.

There are no known side effects associated with this type of ultrasound examination, and it causes no physical discomforts. It is possible that you may feel some slight discomfort from remaining seated for a long period of time. Ultrasound involves no radiation exposure. Although there is no proof of risks of diagnostic ultrasound in pregnant women, pregnant women may wish to exclude themselves from this study to avoid the discomfort of being seated for an extended period.

You are free to stop your participation if you become uncomfortable in any way.

**Benefits**

There is no direct benefit to you for participating in this research. You may benefit indirectly by helping increase knowledge of language, which may have clinical benefits related to speech, or other benefits to society.

**Confidentiality / Data Conservation**

The information in the study records will be kept confidential to the full extent allowed by law. Data will be stored securely on a password-protected secure server, in locked filing cabinets, and behind locked doors at the NCSU Phonology Lab, as well as on a password-protected secure server associated with the Sound Patterns Laboratory at the University of Ottawa. No reference will be made in oral or written reports which could link you to the study. You will NOT be asked to write your name on any study materials so that no one can match your identity to the answers that you provide. If you are video recorded, your identity will be concealed by not showing your eyes or by covering them with eyewear. Data will be retained indefinitely. If you decide to withdraw from the study at any time following data collection, all data collected from you will be destroyed.

**Compensation**

For participating in this study you will receive  \$12.50 (60 minutes) /  \$15 (75 minutes). If you withdraw from the study prior to its completion, you will receive the same compensation. NOTE: If you are an NCSU employee, you will be asked to provide your employment information, for tax purposes.

**What if you have questions about your rights as a research participant?**

If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, by phone: (613) 562-5387, or by email: ethics@uOttawa.ca.

**Consent To Participate**

“I have read and understand the above information. I have received a copy of this form. I agree to participate in this study with the understanding that I may choose not to participate or to stop participating at any time without penalty or loss of benefits to which I am otherwise entitled.”

Subject's signature \_\_\_\_\_ Date \_\_\_\_\_

Investigator's signature \_\_\_\_\_ Date \_\_\_\_\_

## B.2 Language Background Questionnaire

| Basic information                |   |
|----------------------------------|---|
| Code (assigned by researcher):   | <input type="text"/>                                      |
| Year of birth:                   | <input type="text"/>                                      |
| Sex:                             | female <input type="button" value="v"/>                   |
| What is your level of education? | completed primary school <input type="button" value="v"/> |
| Major or field of interest:      | <input type="text"/>                                      |

| Where have you lived, and how old were you at the time? (use as many rows as you need) |                      |                      |                      |                      |
|--|----------------------|----------------------|----------------------|----------------------|
|  | City                 | State/Province       | Country              | Ages                 |
| place #1:  | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| place #2:  | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| place #3:  | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| place #4:  | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| place #5:  | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| Others (describe):   | <input type="text"/> |                      |                      |                      |

| Tell us about your parents or caretakers |  |                      |                               |                      |   |
|--|--|----------------------|-------------------------------|----------------------|---|
|  | Relationship (e.g., mother, foster parent) | Birthplace           | Places lived during childhood | Occupation           | Level of education  |
| caretaker #1:                            | <input type="text"/>                       | <input type="text"/> | <input type="text"/>          | <input type="text"/> | completed primary school <input type="button" value="v"/> |
| caretaker #2:                            | <input type="text"/>                       | <input type="text"/> | <input type="text"/>          | <input type="text"/> | completed primary school <input type="button" value="v"/> |
| Others (describe):                       | <input type="text"/>                       |                      |                               |                      |   |

| What languages do you know? (use as many rows as you need) |                      |                      |   |   |   |   |
|--|----------------------|----------------------|---|---|---|---|
|  | Language             | Age started          | Reading                                     | Writing                                     | Speaking                                    | Comprehension                               |
| Language #1  | <input type="text"/> | <input type="text"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> |
| Language #2  | <input type="text"/> | <input type="text"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> |
| Language #3  | <input type="text"/> | <input type="text"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> |
| Language #4  | <input type="text"/> | <input type="text"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> |
| Language #5  | <input type="text"/> | <input type="text"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> | not at all <input type="button" value="v"/> |
| Notes:   | <input type="text"/> |                      |   |   |   |   |

|   |  |
|---|--|
| What is your dominant hand?                               | right <input type="button" value="v"/> |
| Please describe any speech or hearing disorders you have: | <input type="text"/>                   |

# Bibliography

- Bailey, Guy, Tom Wikle, Jan Tillery, and Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3:241–264.
- Baker, Adam. 2005. *Palatoglossatron 1.0*. University of Arizona, Tucson, Arizona.
- Baker, Adam, Diana Archangeli, and Jeff Mielke. 2011. Variability in English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23:347–374.
- Bass, Michael. 2009. Street or shtreet? Investigating (str-) palatalisation in Colchester English. *Estro: Essex Student Research Online* 10–21.
- Beddor, P. S., and R. A. Krakow. 1999. Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America* 106:2868–2887.
- Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85:785–821.
- Beddor, P.S., K.B. McGowan, J.E. Boland, A.W. Coetzee, and A. Brasher. 2013. The time course of perception of coarticulation. *Journal of the Acoustical Society of America* 133:2350–2366.
- Bell-Berti, Fredericka, Lawrence J Raphael, David B Pisoni, and James R Sawusch. 1979. Some relationships between speech production and perception. *Phonetica* 36:373–383.
- Blevins, Juliette. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.
- Boersma, Paul, and David Weenink. 2007. *Praat: doing phonetics by computer*. URL <http://www.praat.org>, [Computer program].
- Browman, Catherine, and Louis Goldstein. 1995. Gestural syllable position effects in american english. In *Producing speech: Contemporary issues. for katherine safford harris.*, ed. F. R. Bell-Berti and J. Lawrence, 510–520. Woodbury, NY: American Institute of Physics.
- Browman, Catherine P, and Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6:201–251.

- Byrd, Dani. 1993. 54,000 american stops. *UCLA Working Papers in Phonetics* 83:97–116.
- Campbell, Fiona, Bryan Gick, Ian Wilson, and Eric Vatikiotis-Bateson. 2010. Spatial and temporal properties of gestures in North American English /r/. *Language and Speech* 53:49–69.
- Cruttenden, Alan. 2014. *Gimson's Pronunciation of English*. Routledge.
- Darlington, Richard B. 1970. Is Kurtosis Really "Peakedness?". *The American Statistician* 24:19–22.
- Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120:407–415.
- Delattre, Pierre, and Donald C. Freeman. 1968. A dialect study of American r's by x-ray motion picture. *Linguistics* 6:29–68.
- Dodsworth, Robin, and Mary Kohn. 2012. Urban rejection of the vernacular: The SVS undone. *Language Variation and Change* 24:221–245.
- Durian, David. 2007. Getting [ʃ]tronger every day?: More on urbanization and the socio-geographic diffusion of (str) in Columbus, OH. *University of Pennsylvania Working Papers in Linguistics* 13:6.
- Farnetani, Edda, and Daniel Recasens. 2010. Coarticulation and connected speech processes. *The Handbook of Phonetic Sciences, Second Edition* 316–352.
- Fowler, Carol A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8:113–133.
- Fowler, Carol A, and Elliot Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36:171–195.
- Frisch, Stefan A, and Sylvie M Wodzinski. 2016. Velar–vowel coarticulation in a virtual target model of stop production. *Journal of phonetics* 56:52–65.
- Fruehwald, Josef. 2016. The early influence of phonology on a phonetic change. *Language* 92:376–410.
- Galantucci, Bruno, Carol A Fowler, and Michael T Turvey. 2006. The motor theory of speech perception reviewed. *Psychonomic bulletin & review* 13:361–377.
- Gauchat, Louis. 1905. *L'unité phonétique dans le patois d'une commune*. Niemeyer.
- Gick, Bryan. 1999. A gesture-based account of intrusive consonants in english. *Phonology* 16:29–54.

- Gick, Bryan, and Fiona Campbell. 2003. Intergestural timing in english /r/. In *Proceedings of the XVth International Congress of Phonetic Sciences*.
- Gick, Bryan, Fiona Campbell, Sunyoung Oh, and Linda Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34:49–72.
- Gick, Bryan, Ian Wilson, and Donald Derrick. 2013. *Articulatory phonetics*. John Wiley & Sons.
- Glain, Olivier. 2014. Introducing Contemporary Palatalisation. In *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York Conference*, ed. Theodora Lee. York.
- Gordon, Elizabeth, and Margaret Maclagan. 2004. Regional and social differences in new zealand: phonology. In *A handbook of varieties of english*, ed. Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie, and Clive Upton, volume 1, 603–613. Mouton de Gruyter.
- Gu, Chong. 2002. *Smoothing spline ANOVA models*. New York: Springer-Verlag. Springer Series in Statistics.
- Guenther, F., C. Espy-Wilson, S. Boyce, M. Matthies, M. Zandipour, and J. Perkell. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* 105 (5):2854–2865.
- Guenther, F. H. 1994. A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics* 72:43–53.
- Gylfadottir, Duna. 2015. Shtreets of Philadelphia: An acoustic study of /str/-retraction in a naturalistic speech corpus. *University of Pennsylvania Working Papers in Linguistics* 21:89–97.
- Hammarberg, Robert. 1976. The metaphysics of coarticulation. *Journal of Phonetics* 4:353–363.
- Hay, Jennifer, Paul Warren, and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34:458–484.
- Hueber, T., G. Chollet, B. Denby, and M. Stone. 2008. Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application. In *Proceedings of ISSP2008*, 365–369. Strasbourg, France: ISSP.
- Hyman, Larry. 1977. Phonologization. In *Linguistic studies presented to Joseph H. Greenberg*, ed. A. Juilland, 407–418. Saratoga: Anna Libri.
- Hyman, Larry M. 2013. Enlarging the scope of phonologization. *Origins of sound change: Approaches to phonologization* 3–28.

- Jakobson, Roman. 1931. Prinzipien der historischen phonologie. *Travaux du Cercle Linguistique de Prague* 4:247–267.
- Janda, Richard D. 1999. Accounts of phonemic split have been exaggerated - but not enough. In *Proceedings of ICPHS 14*.
- Janda, Richard D, and Brian D Joseph. 2003. Reconsidering the canons of sound-change. In *Historical Linguistics 2001: Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne, 13-17 August 2001*, volume 237, 205. John Benjamins Publishing.
- Jones, Daniel. 1963. *The Pronunciation of English*. Cambridge University Press, fourth edition, revised and enlarged edition.
- Joos, Martin. 1948. Acoustic Phonetics. *Language* 24:5–136.
- Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language variation and change* 2:205–254.
- Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell.
- Lawrence, W. 2000. /str/ → /ʃtr/: Assimilation at a distance? *American Speech* 75.1:82–87.
- Liberman, Alvin M, and Ignatius G Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21:1–36.
- Lindau, Mona. 1985. The story of /r/. In *Phonetic linguistics: Essays in honor of Peter Ladefoged*, ed. Victoria A. Fromkin, 157–168. Orlando, FL: Academic Press.
- Lindblom, B., S. Guion, S. Hura, S.-J. Moon, and R. Willerman. 1995. Is sound change adaptive? *Rivista di Linguistica* 7:5–37.
- Lindblom, Björn. 1983. Economy of speech gestures. In *The production of speech*, ed. MacNeilage. New York: Springer Verlag.
- Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H and H Theory. In *Speech production and speech modelling*, ed. W.J. Hardcastle and A. Marchal, 403–439. Kluwer, Dordrecht.
- Löfqvist, Anders. 1990. Speech as audible gestures. In *Speech production and speech modelling*, 289–322. Springer.
- Magloughlin, Lyra. 2016. Accounting for variability in North American English /ɪ/: Evidence from children’s articulation. *Journal of Phonetics* 54:51–67.
- Mann, Virginia A, and Bruno H Repp. 1980. Influence of vocalic context on perception of the S-s distinction. *Attention, Perception, & Psychophysics* 28:213–228.

- Massaro, D. W., and M. M. Cohen. 1983. Phonological constraints in speech perception. *Perception & Psychophysics* 34:338–348.
- Mathôt, Sebastiaan, Daniel Schreij, and Jan Theeuwes. 2012. Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods* 44:314–324.
- McGowan, Richard S., Susan Nittrouer, and Carol J. Manning. 2004. Development of /[ɪ]/ in young, Midwestern, American children. *Journal of the Acoustical Society of America* 115:871–884.
- Menzerath, Paul, and Armando de Lacerda. 1933. *Koartikulation, steuerung und lautabgrenzung: Eine experimentelle untersuchung*, volume 1. F. Dümmler.
- Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America* 137:2858–2869.
- Mielke, Jeff. 2016. Head-stabilization for ultrasound recording: head-orientation tracking. Technical report, North Carolina State University.
- Mielke, Jeff. 2017. Functions for ssanova comparisons of tongue traces in polar coordinates using gss.
- Mielke, Jeff, Adam Baker, and Diana Archangeli. 2010. Variability and homogeneity in American English /ɪ/ allophony and /s/ retraction. In *Variation, Detail, and Representation (LabPhon 10)*, ed. Cécile Fougeron, Barbara Kühnert, Mariapaola D’Imperio, and Nathalie Vallée, 699–719. Berlin: Mouton de Gruyter.
- Mielke, Jeff, Adam Baker, and Diana Archangeli. 2016. Individual-level contact limits phonological complexity: evidence from bunched and retroflex ɪ. *Language* 92.
- Mielke, Jeff, Christopher Carignan, and Erik R. Thomas. 2017a. The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: an ultrasound study. *Journal of the Acoustical Society of America* 142:332–349.
- Mielke, Jeff, Bridget Smith, and Michael J Fox. 2017b. Implications of covert articulatory variation for several phonetic variables in raleigh, north carolina english. *The Journal of the Acoustical Society of America* 141:3981–3981.
- Miller, George A., and Patricia E. Nicely. 1955. An analysis of perceptual confusions among some english consonants. *Journal of the Acoustical Society of America* 27:338–352.
- Ohala, John J. 1981. The listener as a source of sound change. In *CLS 17: papers from the parasession on language and behavior*, ed. C.S. Masek, R.A. Hendrik, and M.F. Miller, 178–203. Chicago: CLS.

- Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, ed. MacNeilage, chapter 9, 189–216. New York: Springer.
- Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In *Language change: Contributions to the study of its causes (Trends in linguistics, studies and monographs no. 43)*, ed. L. E. Breivik and E. H. Jahr, 173–198. Berlin: Mouton de Gruyter.
- Ohala, John J. 1993. Coarticulation and phonology. *Language and speech* 36:155–170.
- Ohala, John J. 1994. Hierarchies of environments for sound variation; plus implications for ‘neutral’ vowels in vowel harmony. *Acta Linguistica Hafniensia* 27:371–382.
- Ohala, John J, and Maria-Josep Solé. 2010. Turbulence and phonology. *Turbulent sounds: an interdisciplinary guide* 37–102.
- Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39:151–168.
- Olive, Joseph P, Alice Greenwood, and John Coleman. 1993. *Acoustics of american english speech: a dynamic approach*. Springer Science & Business Media.
- Paul, Hermann. 1880. *Prinzipien der sprachgeschichte*. Halle: Niemeyer.
- Peirce, Jonathan W. 2007. Psychopy—psychophysics software in python. *Journal of neuroscience methods* 162:8–13.
- Pierrehumbert, Janet B. 2002. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the emergence of linguistic structure*, ed. Joan Bybee and Paul Hopper, 323–418. Amsterdam: John Benjamins.
- Pitt, Mark A. 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & psychophysics* 60:941–951.
- R Development Core Team. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>, ISBN 3-900051-07-0.
- Read, Charles. 1971. Pre-school children’s knowledge of English phonology. *Harvard educational review* 41:1–34.
- Read, Charles, and Rebecca Treiman. 2012. Children’s invented spelling: What we have learned in forty years. *Rich Languages from Poor Inputs* 195–209.
- Recasens, Daniel, Maria Dolors Pallarès, and Jordi Fontdevila. 1997. A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America* 102:544–561.

- Rosenbaum, David A, and Matthew J Jorgensen. 1992. Planning macroscopic aspects of manual control. *Human Movement Science* 11:61–69.
- Rousselot, P.J. 1901. *Principes de phonétique expérimentale*, volume 1. Paris: H. Welter.
- Rutter, Ben. 2011. Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English. *Journal of the International Phonetic Association* 41:27–40.
- Saltzman, Elliot L, and Kevin G Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological psychology* 1:333–382.
- Sander, E.K. 1972. When are speech sounds learned? *Journal of Speech and Hearing Disorders* 37:55–63.
- Shapiro, M. 1995. A case of distant assimilation: /str/ → /ʃtr/. *American Speech* 70.1:101–107.
- Shultz, A.A., A.L. Francis, and F. Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America* 132:EL95–EL101.
- Smit, Ann Bosma. 1993. Phonological error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech and Hearing Research* 36:533–547.
- Smit, Ann Bosma, Linda Hand, J. Joseph Freilinger, John E. Bernthal, and Ann Bird. 1990. The Iowa articulation norms and its Nebraska replication. *Journal of Speech and Hearing Disorders* 55:779–798.
- Smith, Bridget. 2013. The Interaction of Speech Perception and Production in Laboratory Sound Change. Doctoral Dissertation, Ohio State University.
- Smits, R. 2001. Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance* 27.
- Sproat, Richard, and Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21:291–311.
- Stetson, R. 1951. *Motor phonetics: A study of speech movements in action*. Amsterdam: North-Holland Publishing Company, 2nd edition.
- Stevens, Kenneth N. 1998. *Acoustic Phonetics*, volume 30. Cambridge, MA.
- Stevens, Mary, and Jonathan Harrington. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58:118 – 134.

- Stoel-Gammon, Carol. 1985. Phonetic inventories, 15–24 months: a longitudinal study. *Journal of Speech, Language, and Hearing Research* 28:505–512.
- Stone, Maureen. 2005. A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics* 19:455–502.
- Sweet, Henry. 1877. *A Handbook of Phonetics*, volume 2. Clarendon Press.
- Thomas, Erik R. 2011. *Sociophonetics: An introduction*. New York: Palgrave Macmillan.
- Twist, Alina, Adam Baker, Jeff Mielke, and Diana Archangeli. 2007. Are ‘covert’ /ɪ/ allophones really indistinguishable? *Penn Working Papers in Linguistics* 13.2.
- Van Rossum, Guido, et al. 2007. Python programming language. In *USENIX Annual Technical Conference*, volume 41, 36.
- Walker, Rachel, Michael Proctor, Caitlin Smith, and Ewald Enzinger. 2016. Asymmetries in English Liquid Production and Vowel Interactions. In *15th Conference on Laboratory Phonology*. Ithaca, NY.
- Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics*, ed. W. Lehmann and Y. Malkiel. Austin: U. of Texas Press.
- Westbury, J., M. Hashi, and M. Lindstrom. 1998. Differences among speakers in lingual articulation for American English /ɪ/. *Speech Communication* 26:203–226.
- Whalen, D. H. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]-[S] boundary. *Journal of the Acoustical Society of America* 69:275–282.
- Whitney, William Dwight. 1867. *Language and the study of language: Twelve lectures on the principles of linguistic science*. N. Trübner & Company.
- Wilbanks, Eric. 2017. Social and Structural Constraints on a Phonetically-motivated Change in Progress: (str) Retraction in Raleigh, NC. *University of Pennsylvania Working Papers in Linguistics* 23:33.
- Wilbanks, Eric, and Jeff Mielke. 2015. One script to rule them all. Praat script, NCSU Phonology Lab.
- Young, Steve, Gunnar Evermann, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Valtcho Valtchev, and Phil Woodland. 2002. *HTKBook*. Cambridge University Engineering Department.
- Yuan, Jiahong, and Mark Liberman. 2008. Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America* 123:3878.

Yuan, Jiahong, and Mark Liberman. 2009. Investigating /l/ variation in English through forced alignment. In *Tenth Annual Conference of the International Speech Communication Association*.