A Sociophonetic Analysis of New York City English in the Speech of Teenagers from Nassau County, Long Island

by

Allison Shapp

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

New York University

January 2019

_____

John Victor Singler

DEDICATION

For Mom and Dad

ACKNOWLEDGEMENTS

There are so many people besides me who deserve credit for this accomplishment. I will attempt to acknowledge and thank as many of them as I can here, but please forgive my dissertation-addled brain if I forget anyone.

I must start with my advisor John Singler. Thank you for always steering me in the right direction, for the snarky and (usually) amusing comments on my writing, and for all the two a.m. email exchanges. You have always managed to be both kind and demanding, which turned out to be exactly what I needed. I am honored to help you close out an illustrious career of raising students into successful scholars, though I imagine that the other aspects of your research career will not be ending *any* time soon.

To the rest of my committee: Renée Blake, thank you for the opportunities you have given me to work with you on such interesting and wide-ranging projects, as well as for periodically reminding me that I am a person and not just a student. Greg Guy, thank you for starting me off with a solid foundation in variationist sociolinguistics and also for always being so excited to discuss short-*a* with me. To Lisa Davidson and Cece Cutler, thank you both for your input and guidance on my dissertation, and most of all for your patience.

I would like to acknowledge my fellow graduate students, especially those in the NYU Sociolab, for being such a supportive group of people. I have learned so much from each of you. Some of the many friends that I have met on this journey are: Isaac Bleaman, Marie-Eve Bouchard, Carina Bauman, Libby Coggshall, Dan Duncan, Zack Jaggers, Nicole Holliday, Sonia

Kasyanenko, Jeremy Kuhn, Tal Linzen, Sean Martin, Luiza Newlin-Łukowicz, Natalie Povilonis de Vilchez, Cara Shousterman, and Dani Szeredi. Amy Wong and Emily Nguyen also belong on this list, but they get special shout-outs for so generously lending me their audio equipment with which to conduct my research.

I want to especially thank Nathan LaFave for being my partner in crime. Without your camaraderie and friendship, graduate school would have been a very different, and much less bearable, experience.

An important thank you to Teresa and Hannah, without whom everything would truly be chaos. Thank you to Miriam Kates and my intrepid band of undergraduate transcribers: you saved me so many hours of my life, not to mention my sanity. Thank you also to Sophia Snyder, Ben Williams, Tim Zirkel, and Kate Mooney for making yourselves available for last-minute proofreading assistance. Any remaining mistakes are solely my own responsibility.

This work was made possible by a GSAS Doctoral Dissertation Fellowship, an NYU Global Research Initiative grant, and a Henry M. MacCracken Fellowship from NYU. I am also indebted to FIRST Robotics and the teachers, mentors, and students there who welcomed me into the fold. I thank these organizations for their generous support.

I would like to say a very special thank you to Anna Marie Trester, who is just an inspiration of a person. Since the very first time I met you, you have been so generous with your time and wisdom. I am so grateful that you have become a mentor and a real friend. Your CL network has been an immeasurable source of support for me during this process, and without you bringing the magic of The Porches into my life, I might still be stuck. From the bottom of my heart, thank you for everything.

To Wellspring House and Trudy at The Porches, thank you for welcoming me as a writer-in-residence and for providing me peaceful spaces in which to fully focus on finishing this work. Both of these writing retreats were transformative experiences. In the same vein, thank you also to Nellie for your understanding and guidance throughout the best and worst (mostly worst) of grad school.

I am so fortunate to have such a caring network of people, both friends and family, who support me unconditionally. You have all played important roles in this achievement, many of you as invaluable (and sometimes involuntary) on-call dialect consultants! An especially big thank you to Lisa, Miranda, and Suzie for helping keep my heart full and my mind somewhat sane. Thank you Honey Nor, for fostering my love of words, even if you did always make me go look everything up myself in the Funk & Wagnalls. I am truly sorry that I can't name everyone, but to all the other members of my extended family who have followed my academic journey with excitement: thank you.

Brian and Jenny, you have been the best siblings with which to share a childhood and to actually still enjoy the accompany of as adults! Thank you for all your support, and for all the goofing off and extreme indulgence in our shared nerdiness.

Finally, to my parents, thank you for instilling in me a love of knowledge and a drive to succeed. I would not be who I am today without your love, your support, and your belief in me even when I didn't believe in myself. This is as much your accomplishment as it is mine.

ABSTRACT

This study examines the ways in which recent changes in the English spoken in New York City (NYC) are reflected in the speech of teenagers in a suburban area of the NYC dialect region: Nassau County, Long Island. To address this question, quantitative and instrumental sociophonetic techniques are employed to analyze two phonological features of the dialect: the use of a THOUGHT vowel that is raised and distinct from the LOT vowel, and the short-*a* split. Both of these variables have been well documented as features of NYC English for over a century but have been shown in recent work (e.g. Becker 2010) to be disappearing from the repertoire of young speakers in NYC.

The community of study is 24 students (aged 14-18) who are members of their high school's FIRST Robotics team, as well as seven mentors of the same team who range in age from 24 to 86 years old. The researcher spent the 2015-2016 school year acting as a mentor of the team herself, while carrying out participant observation and ethnographic research in the community. The data comes from sociolinguistic interviews conducted with each of the 31 participants between April and June of 2016.

The results show that all of the speakers in the study have significantly separate LOT and THOUGHT vowel classes, demonstrating no signs of imminent merger. Students all produced at least some THOUGHT-raising, and several used the feature to a large extent. Comparing the mentors to the students, however, does show an overall decrease of THOUGHT-raising in apparent time. Analysis of the students' short-*a* systems found that four maintain a traditional NYC

complex short-*a* split, and four have a transitional system which includes *tensed* short-*a* in words of the BAD word class, but not in words of the BASH word class. The remaining sixteen students were found to have nasal short-*a* splits, but half of them also maintain an Open Syllable Constraint, representing an additional transitional phase between the NYC complex split and the more common nasal split. Further, a correlation is found between rates of THOUGHT-raising and conservation of the NYC complex short-*a* split. This suggests that there is some parallel development of these two vowels in the speech of Long Islanders.

Recent work in the region has proposed that the complex NYC short-*a* split is being lost, as well as suggesting the presence of an apparent time change toward reduction of THOUGHT-raising. The current study suggests that these language changes are taking a different, or at least a slower, path in parts of the dialect region outside of the core area of Manhattan. The results also suggest a connection between higher rates of use of the New York City variants and a stronger affiliation by the speaker with a local, Long Island identity. The study concludes that there is a possibility of phonological features that have long been associated with a "New York City accent" moving instead towards the indexing of a Long Island identity and dialect.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

Chapter 1: Introduction

Like many sociolinguistic researchers, I chose a topic to investigate that involves aspects of my own social and linguistic identity. My family immigrated to New York City from Eastern Europe three generations ago and followed a trajectory common to many families who ended up in the suburbs of Long Island. My grandparents were born and lived in New York City, while my parents were both born in the city, moved out to Long Island as children, and raised in the suburbs. Despite my ancestral connection to New York City, and living in such close proximity to it for my whole life, there has always been a sense that the identity of being a "New Yorker" was elusive, elite, and not something I was allowed to claim.

It was not until I joined the linguistics department at NYU that I started to evaluate what it meant to be a New Yorker in a new way, in terms of language. I entered the department during something of a boom in sociolinguistic research about New York City, and was immediately drawn in. For much of my time in the department, I have been the only "native New Yorker," in terms of being the person other students ask to consult with on matters of local dialect and to record experimental stimuli for them with a seemingly masterful control of a myriad of vowels. But even just writing the term "native New Yorker" in the previous sentence made me uncomfortable. I'm from Long Island, I can't say that. Linguistically, however, I found that this categorization did not seem to hold. Everything I was learning about the history of New York City English (NYCE) was familiar. When I read the recent work that showed these NYCE dialect features like the short-*a* split, THOUGHT-raising, and /r/-vocalization disappearing from

the local dialect, and seemingly completely absent from some of the youngest speakers in the area, it did not line up with my own experience and what I heard growing up in Nassau County. I read some of the older descriptions of the short-*a* split and was amazed by how well it described my own intuitions. I hypothesized that this decline in NYC features was either not happening in Nassau County, or at least was not as progressed. In the following chapter, I do not propose that the decline of these features is not occurring on Long Island at all; for example, /r/-vocalization has all but completed its disappearance in my generation. What I *will* show is that my hypotheses about short-*a* and the THOUGHT-vowel were borne out. While productions of both vowels are trending away from their historical New York City values and towards the national variants, they are still present in even young speakers in Nassau County, and with a much higher frequency and degree than would be predicted by the previous work on these features that has largely been situated in Manhattan.

1.1 The Questions

This dissertation seeks to investigate the state of the New York City dialect of English as it exists currently in Nassau County, particularly among young speakers. By looking in-depth at THOUGHT-raising and the short-*a* split, I ask if these features are more prevalent in Nassau County than among young speakers in Manhattan. I also investigate which social factors might affect the degree to which Long Islanders use features of NYCE and the manner in which they deploy them. Finally, through this sociophonetic analysis, I seek to better understand how teenagers on Long Island feel about their identity as "New Yorkers," and if they feel a

connection with New York City. I investigate what connections, if any, these attitudes might have on Long Island speakers' use of phonological features traditionally associated with New York City.

1.2 NYC as a Center of Dialectological Study

New York City has always been an important site for the study of dialectology, described by Babbitt in 1896, Kurath and McDavid in 1961, and prodigiously by William Labov. Labov's 1966 study, *The Social Stratification of English in New York City*, set forward the principles and methodology that established the field of variationist sociolinguistics.

Recently, a resurgence of work on New York City English has been underway, by researchers such as Becker (2010), who revisited the Lower East Side, and Wong (2010, 2015), who investigated the use of NYC features in the English of Chinese-Americans[1]. Becker and Wong's work in Manhattan suggests that there have been changes in the status of the classic variables mentioned in earlier work, indicating that younger speakers are moving away from them towards General American English. Newman (2014) however, found higher rates of NYC features in his younger speakers in Brooklyn and Queens than did Becker and Wong in Manhattan.

Bronstein 1962 describes the dialect area of NYCE as a "core" area surrounded by an "inner ring" and an "outer ring." For Bronstein, the core dialect area of New York includes four

[1]Also see Blake and Shousterman 2010, Coggshall and Becker 2010, Newlin-Łukowicz 2015, Shousterman 2015.

of the city's five boroughs (Manhattan, the Bronx, Brooklyn, and Queens) as well as Hudson County, New Jersey. The "inner ring" includes three New York counties, Nassau County among them. Bronstein includes Suffolk County in the "outer ring" of NYCE, which he does not consider part of the NYC dialect area. Past and recent work has largely focused on Manhattan as the true "core" of the dialect region. Little if any linguistic research has been done on the NYC inner ring. The present research seeks to fill that gap.

1.3 The Variables Under Investigation

This dissertation focuses on two phonological variables of NYCE: short-*a* and THOUGHT-raising. Short-*a* refers to words that have the underlying phoneme /æ/ in General American English (GAE). In GAE the most common pattern for the pronunciation of /æ/ is the "nasal" split, where the vowel is *tense* (raised, fronted) when followed by a nasal consonant and lax (lowered, backed) otherwise. In contrast, historically New York City English has had a "complex short-*a* split." Here the basis for the split, while primarily the following environment, is not nasal vs. oral but is instead front nasals, voiced stops, voiceless fricatives (*tense*) vs. velar nasals, voiceless stops, and voiced fricatives (*lax*). When the vowel is word-initial, in an open syllable, or in most function words, it is lax regardless of following environment. A similar pattern exists in Philadelphia, but it is not exactly the same as that found in New York. There is also a pattern emerging in the United States in areas affected by the Northern Cities Shift in which all short-*a* is raised and ingliding.

THOUGHT-raising is the pronunciation of the THOUGHT vowel with a lowered F1, generally described as below 700Hz. In Labov 1966, a progressive increase in THOUGHT-raising in New York City English is presented as a change in progress from below, with the frequency and extent to which the vowel is raised increasing over time. Becker (2010) proposes that since then, the direction of the change has reversed, and that there is now a change from above resulting in reduced THOUGHT-raising.

Both of these variables show changes over time and are reported in Becker 2010, *inter alia,* to not be as prevalent among young speakers in Manhattan as they used to be. Newman (2014) shows that the traditional pronunciations of these vowels holds to a higher degree in the outer boroughs of New York City than it does in Manhattan. This dissertation explores what the status of these New York City phonological variables are among young speakers in Nassau County, the suburban area of Long Island that borders the borough of Queens.

1.6 Road Map

In the following chapters, I present data gathered during a year of field work with teenagers at a high school in Nassau County on Long Island. In Chapter 2, I describe the history of Long Island and how the research site, the town of Antioch, fits into the area's social history. In Chapter 3, I outline the methodology I followed and introduce the participants in my study. Chapter 4 gives an in-depth outline of previous research on short-*a* in New York City English, spanning more than a century, before I present my own results and interpretation in Chapter 5. Chapter 6 presents a quantitative description of THOUGHT-raising for this sample of speakers.

Chapter 7 concludes by summarizing my results, connecting them to matters of identity, and suggesting future directions in which research like this may develop.

Chapter 2: Long Island, Antioch, Grumman High School and the CyberPilots

2.1 Introduction

The purpose of this chapter is to give an overview of the research site, Grumman High School (GHS), as well as the town in which the school is located: Antioch, New York[2]. The town of Antioch is found on Long Island, and in section 2.2, I first give a description of the geography and demographic history of Long Island in order to contextualize and situate Antioch and its residents in their geographical surroundings. I also give a brief overview of the history of linguistic and dialectal study of Long Island and the New York City area, in order to explain my rationale for studying a town in Nassau County like Antioch. Finally, I give a description of the Community of Practice at hand: the CyberPilots, GHS's extracurricular robotics team, which is part of the national FIRST Robotics organization.

2.2 Long Island

Long Island is a 118-mile long island off the coast of the southern tip of the State of New York, just east of Manhattan. It is the largest island adjoining the continental United States. It is

[2]The names of the town (Antioch), the school (Grumman High School), and the robotics team (the CyberPilots) are all pseudonyms.

twenty miles across, north to south, at its widest point. To the north of the island is the Long Island Sound, and to the south and east is the Atlantic Ocean.

2.2.1 Early Settlement

Long Island was colonized by both the Dutch and the English in the early seventeenth century. In 1621, The Dutch West India Company established the colony of New Netherlands (Charter of the Dutch West India Company 1621), which grew to encompass all of present-day New York City as well as Nassau County (and parts of Connecticut and New Jersey). The British gained a foothold on Long Island in the 1640s as groups of settlers from Connecticut traveled south across the Sound, first claiming land on the east end of the island and then encroaching further west. The British ultimately won the province of New Netherlands from the Dutch in the Second Anglo-Dutch War of 1664 and renamed it New York. As of 1683, Long Island was divided into three counties: Kings, Queens, and Suffolk. Queens County encompassed western Long Island, including the present-day Nassau towns of Hempstead and Oyster Bay.

On January 1st, 1898, all of the western towns in Queens County became part of New York City. The eastern towns (Hempstead, North Hempstead, and Oyster Bay) were not made part of Greater New York but still remained part of Queens County. During the following year, these three towns pushed for secession from Queens and for the creation of a new county. The name “Nassau” was proposed to reflect the region’s early settlement by both the Dutch and the English, in that King William III of the House of Orange-Nassau was a monarch of both England and the Netherlands. The creation of Nassau County took effect on January 1, 1899.

Presently, Long Island consists of four counties: Queens and Kings (Brooklyn) Counties, which are boroughs of the city of New York, and Nassau and Suffolk Counties, which are not. Because of this political division at the city limits, the convention is that when using the term "Long Island," one is referring to only Nassau and Suffolk Counties.

Figure 1: Map of Long Island Counties

2.2.2 The Growth of Nassau County

Throughout the nineteenth century, Long Island was mainly rural and agricultural. The Long Island Railroad (LIRR) was established in 1836, opening up routes for easier transportation and shipment of goods from New York City as well as transportation of the produce grown on Long Island to the markets of New York City. In 1910, the LIRR established direct rail service to Pennsylvania Station in midtown Manhattan (Seyfried 1984:74). This allowed for the population of the previously small villages in Nassau County to swell with commuters over the next decades, followed by businesses to support them. It is now the busiest commuter railway in America (American Public Transportation Association 2016). The population growth of Nassau

County from its inception in 1900 until 2014 are shown in Figures 2 and 3 below (the data in these figures was compiled by Population.us).

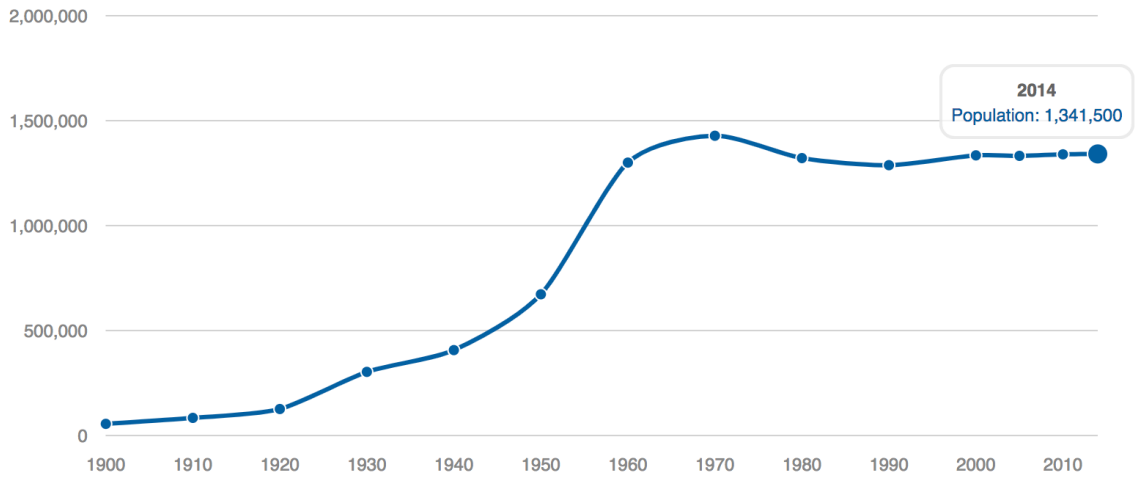


Figure 2: Population Growth in Nassau County from 1900 to 2014

Annual growth rate

[1900-1910] **+4.23** %/yr
[1910-1920] **+4.16** %/yr
[1920-1930] **+9.16** %/yr
[1930-1940] **+2.99** %/yr
[1940-1950] **+5.16** %/yr
[1950-1960] **+6.81** %/yr
[1960-1970] **+0.94** %/yr
[1970-1980] **-0.77** %/yr
[1980-1990] **-0.26** %/yr
[1990-2000] **+0.36** %/yr
[2000-2005] **-0.03** %/yr
[2005-2010] **+0.11** %/yr
[2010-2014] **+0.04** %/yr

Figure 3: Growth Rates of the Population of Nassau County by Decade

The figures above show that the population of Nassau County grew continuously towards a peak in 1970, before slightly decreasing and somewhat leveling off at the current population of around 1.3 million people. The steepest periods of growth were between 1920-1930 (9.16%) and between 1950-1960 (5.81% growth). The 1920 boom was a consequence of the above-mentioned establishment of a LIRR route into New York's Penn Station. The second wave of settlement occurred immediately after WWII, when returning soldiers were offered government incentives to move their families into new housing developments like those built by William Levitt, making Nassau County the nation's largest suburban area at the time. This is the era when most of the families of my study's participants settled on Long Island.

This period also saw a boom in aerospace engineering in the United States, which Long Island played a central part in. Nassau County was home to many of the largest aerospace engineering companies such as Grumman and Republic. Long Island has been dubbed the Cradle of Aviation because of its history as a site from which many early aviators took off for their attempts at solo flights across the Atlantic Ocean. Roosevelt Field, located in Nassau County, was the departure point of many well-known pilots of the twentieth century, such as Charles Lindbergh and Amelia Earhart. During World War II, Long Island companies built 46% of American fighter planes (Cradle of Aviation Museum, n.d.). This established aviation industry allowed Long Island to continue to prosper during the transition from aviation to aerospace engineering, leading to the Lunar Module being constructed at the Grumman factory in Hicksville in the 1960s.

Present day Nassau County has an area of 287 square miles and an estimated population in 2017 of 1,369,514 (US Census). It is the most densely populated county in New York State outside of New York City. The population of Nassau County is 74.4% White, 12.9% Black or African-American, 10.1% Asian, with the remaining 2.5% made up of American Indian, Hawaiian or Pacific Islander, and mixed-race persons. 17.2% of the total population also identifies as Hispanic or Latino. Among residents over the age of 25, 90% have graduated high school, and 43.5% have earned a bachelor's degree or higher (US Census). Nassau County is composed of three towns: Hempstead, North Hempstead, and Oyster Bay. The County Seat is located in Mineola, which is marked with a green star in Figure 4 below.

Figure 4: Map of Nassau County

2.2.3 The Levittown

Figure 5: Photograph of a street in Levittown, New York in the early 1950s

After WWII, William Levitt and his company, Levitt & Sons, created seven developments in the United States that were specifically geared towards returning veterans and their new families. In 1950, *Time* Magazine estimated that Levitt and Sons built one out of every eight houses in United States. The first of these towns was built on Long Island in Nassau County, and was partially supported through government subsidies from the Federal Housing Administration, which was working to alleviate the post-war housing shortage. Levitt houses were built in an efficient assembly-line fashion, with crews of construction workers each trained to perform one specific task in the building of a house. Large numbers of identical houses were

constructed very quickly, and whole houses could be entirely built in one day. There were 1,400 of these Levitt houses purchased during the first three hours of sales on Long Island in March 1947. They cost around $7,000 with no money down required of veterans.

Each house included many of the idealized symbols of suburban life: white picket fences, green lawns, and modern appliances. This idyllic promise of the American Dream did not apply to all races, however. The standard lease agreement signed by early residents of Levittown included a “restrictive covenant” clause that explicitly stated the house could not "be used or occupied by any person other than members of the Caucasian race." (Lambert 1997) This was removed from leases after a Supreme Court Case in 1948 ruled such restrictions to be “unenforceable as law,” but Levitt did not change his personal policy and the precedent had been set for these communities to exclude minorities.

Levittowns have come to represent segregation and white flight from the city, two things that are historically entrenched in the towns and neighborhoods of Long Island, and that still persist to a problematic degree today. But for many, these suburbs were a symbol optimism after an era of wartime. Antioch is an example of this type of suburban development that was built with government subsidies after WWII (though not by Levitt himself), and it still largely reflects both the pleasant nostalgia and the legacy of the harmful racial segregation of the time.

2.3 Previous Linguistic Study

As introduced in Chapter 1, New York City has been an important site for the study and development of the fields of dialectology and sociolinguistics since at least as far back as the late

19th century. Yet, there has been little written about the areas directly surrounding New York City, such as Long Island, and how they should be categorized in regard to dialect (but see Coggshall 2017 for an in-depth discussion of Jersey City, NJ).

Although the exact eastern boundary of the NYC dialect region has never been precisely drawn, Nassau County has always been considered part of the NYC region in dialectological studies and on isogloss maps (Kurath 1949, Labov, Ash, and Boberg 2006) while most of Suffolk County has usually been considered part of New England, an area to which it has been connected historically through the maritime trade (Bakht 2010). In Figure 6, the map from Kurath 1949 shows the isogloss for the Metropolitan New York dialect region (6) cutting straight through the middle of Long Island, basically at the border of the two counties, classifying Nassau County as part of NYC and leaving the status of Suffolk County somewhat ambiguous. This map might be intending Suffolk to be part of the Southwestern New England dialect region, but the isogloss line doesn't make it clear, so the status of the eastern county may just be left unaddressed. There has been scant work done to give better insight as to how Long Island should be classified in regard to these dialect delineations. After fifty years of study of the region, Labov mentions in a footnote that "the eastward line of demarcation in Long Island has not been well defined in any recent studies." (2007:356). Recently there have been two sociolinguistic dissertations done in Suffolk County (Bakht 2010, Olivo 2013) which I briefly describe below, but the current study is the first about Nassau County specifically.

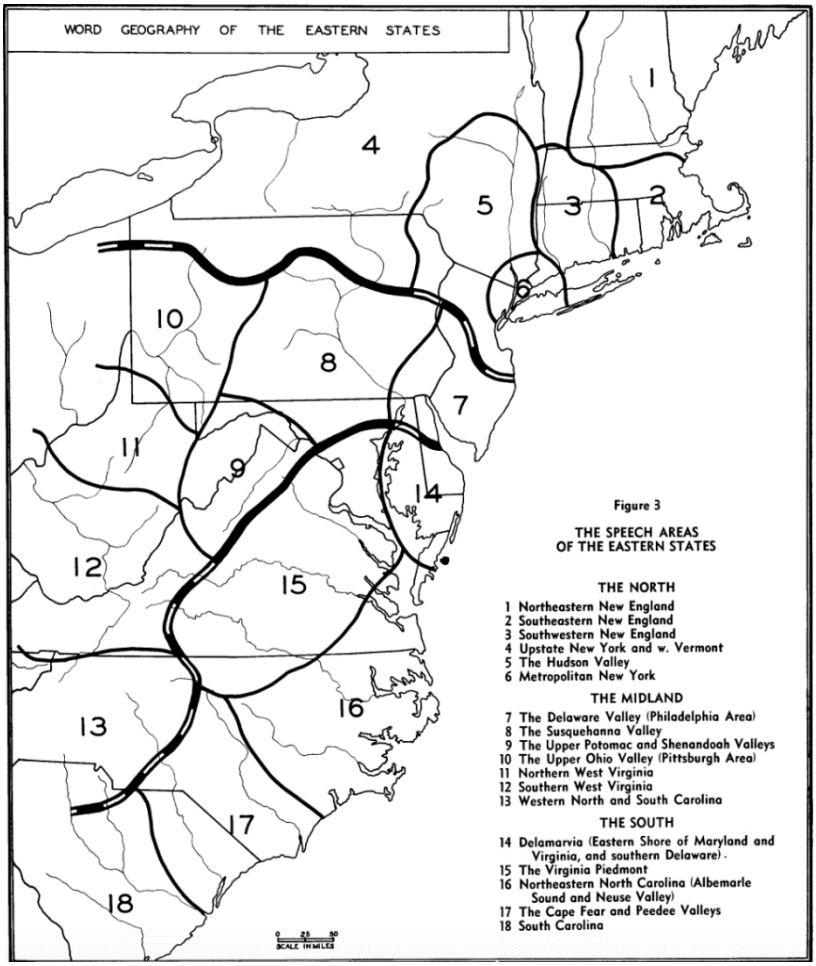


Figure 6: Dialect Regions of the Eastern United States, from Kurath (1949)

Bakht 2010 reports on an extensive ethnographic study of a middle school on the South Fork of eastern Suffolk. Her focus is not on investigating her community's affiliation with a regional dialect, but with features of language that locally index youth culture. She says that, in fact, "the lack of strictly defined dialect boundaries in this area may be a feature that is in itself a factor that contributes to the study of linguistic style" (95). Since the speakers in the area do not have strong regional dialect norms to affiliate with, "the ways that the speakers access and use

linguistic resources is indicative of the linguistic ideologies and social orientation to which the speakers align" (95).

Olivo's 2013 study includes participants from across Suffolk County, who she categorized into demographic groups that she found were locally relevant in discourse analyses of her interviews. These emergent categories included both ethnicity and region, crucially, separating Italian as a locally salient ethnicity not grouped with everyone else of "European Descent." She looked at several phonological features of NYCE in her Long Island speakers who ranged in age from 21 to 84 (37). Across all but one variable, older Long Islanders preferred the traditional NYCE variants (135). She also found that among young speakers, "those who identify with their families' ancestral immigrant pasts tend to prefer the traditional NYCE features," retaining what she calls a "Strong Island" sound to their speech. Olivo says that "by using the phrase *Strong Island*, Long Islanders are establishing themselves as part of the greater NYC metropolitan area, but at the same time recognizing themselves as a separate community" (68). This is a term more strongly associated with the "South Shore" persona described below. Olivo concludes with stating that she thinks New York City has become "one of the first American regions to take on an *ethnic* affiliation" (203).

My study builds on these two previous studies in several ways. I focus on speakers who are in the adolescent peak of language change and are part of one Community of Practice, as Bakht's work does. However, I use quantitative methods more similar to Olivo's in the analysis of phonological variables, finding variation by age, gender, and ethnicity. I focus on just two variables that I analyze in depth, while Olivo did a broad survey analysis of many phonetic features. The current study also adds to the existing literature by being the first study to focus

solely on Nassau County, an area that has always had a close relationship to New York City, sometimes being subsumed within it and sometimes excluded from it, leaving its residents with an ambiguous "New Yorker" identity to navigate.

2.4 Antioch, New York: Location and Demographics

2.4.1 Location

Like much of the area, Antioch was established in the 1950s, before which it was farmland. Towns like Antioch were built during the suburbanization boom after WWII in the Levittown model (see section 2.3.2) and are generally sprawling areas of housing developments and shopping centers with no "downtown" or central area designed into the town.

Long Island has a culturally and historically salient divide between the population of the North Shore and that of the South Shore. The North Shore was settled earlier, and the towns there are smaller, older and nestled into the inlets of the Long Island Sound. The population also tends to be wealthier, illustrated in the extreme by the Gold Coast area depicted most famously in *The Great Gatsby*. The South Shore towns are larger, somewhat newer, and associated with the working class. One can easily find references to this cultural divide in popular media, for example in internet articles with titles like "How you know you're from the South Shore" or "Top 10 Differences Between the North and South Shore." Below is an excerpt from one such article published in the Long Island Press. This sample of items one might find on these lists makes it clear that the North Shore constructed persona is associated with sophistication and perhaps "snootiness," while the South Shore is associated with the working class and perhaps

"trashiness." There are both positive and negative connotations to the characterizations of both groups, and the "rivalry" is almost always light-hearted, intended to give each group an identity to claim and be proud of.

"Top 10 Differences Between the North and South Shores of Long Island"

- Wine from the bottle (NS) vs. Wine from a box (SS).
- Blue blazers and khakis are the style of choice for North Shore males, while jeans and hoodies are the casual go-to for South Shore guys.
- Youngsters brought up on Mozart and Beethoven (NS) vs. rap and rock (SS). But we're all raised on Billy Joel.
- Aged Scotch (NS) vs. Coors Light (SS)

Figure 7: Excerpt from a satirical Long Island Press Article comparing the cultures of the North Shore and South Shore of Long Island

Antioch is in the middle of Long Island in regard to the North/South direction and is largely outside this opposition between the North and South Shore cultures. Many students in Antioch are not aware of this divide if they do not have friends outside of the community. Antioch is also near the border of Nassau and Suffolk County, making it "middle" in all ways.

There is not an analogous divide at the border of Nassau and Suffolk counties. Although Suffolk County spans a relatively much larger portion of the island geographically than does Nassau County, the populations of the two are roughly the same, and the two counties are thought of as having equal stature and claim on the island. At the boundary between the two counties, there is no official or observable landmark to indicate the transition out of one and into the other. The towns blend into each other at this point on the island, and there is no stark divide. The unique characteristics associated specifically with Suffolk County only emerge once you

travel farther east, where the towns start to become more sprawling and eventually turn into farmland and vineyards once you reach the forks of the island. The South Fork is also known for the area called the Hamptons, one of the most expensive and popular seaside resort areas in the country.

2.4.2 Demographics

Antioch has a population of about 26,000 people living within a little less than six square miles. According to the 2016 American Community Survey, the population is 78% White, 14% Asian, 5% Hispanic, and just under 1% Black. The population includes 17.5% who are "foreign born" with "Asian" listed as the majority of foreign-born residents by far. It is important to note that Antioch, and several other towns around it in Nassau County, are heavily Jewish. This is not so uncommon in the New York City area,[3] but it is an unusual demographic makeup compared to the rest of the United States. A 2016 Gallup Poll estimated the US Jewish population at about 2%, while it is estimated to be 14% within the NYC Metropolitan Area. The population of Antioch certainly has a much higher percentage than that, but since the US Census has not asked questions about religion since 1950, it is difficult to find accurate local information about Jewish populations. In 2011 the UJA Federation released the Jewish Community Study: Geographic Profile, which investigated aspects of Jewish populations in eight counties: the five boroughs of New York City, plus Nassau, Suffolk, and Westchester counties. The focus of the study is

[3]There are approximately 1.5 million Jews in the New York City metropolitan area, making it the second largest metropolitan Jewish community in the world, after the Tel Aviv Metropolitan Area in Israel.

primarily levels of poverty in the Jewish community, but it does report some statistics that give insights into the Jewish population of Nassau County, if not individual towns. According to this study, Nassau has the third highest population of Jewish residents of the counties in the study, behind Brooklyn and Manhattan (97). It also mentions that the specific subsection of Nassau where Antioch is located has "the largest proportion of Reform Jews in the eight-county area" (97). The high Jewish population is a defining characteristic of Antioch, and it adds a facet of shared experience among many of the students in my study who grew up attending the Antioch Jewish Center and Hebrew School together.

Many current residents of Antioch share a similar history as to how their families ended up settling on Long Island. For many families, their grandparents or great grandparents immigrated to New York from Europe (mostly eastern Europe, primarily Ashkenazi Jews). For the majority of the students in my study, either their parents or grandparents grew up in New York City and then moved their families to the suburbs of Long Island to raise their children. This is a very common path of migration amongst the families in Antioch and Nassau County in general. This is exactly the story of my own family as well. Many Long Islanders still have family members that live in New York City, and therefore remain closely connected to the city. Many of my participants told stories of visiting their families for the holidays in Brooklyn or elsewhere in New York City.

This immigration pattern does not account for all of the current residents of Antioch, though. There have been more recent waves of immigration to Long Island, most notably from Korea and South East Asia. Nassau County has one of the largest Sikh populations in the United States, with a large central temple located there. In this study, the majority of the speakers are

Jewish, while there is a minority of speakers who are Indian and Chinese. There are also a handful of speakers who represent the other dominant ethno-religious group on Long Island, Italian Catholics.

2.5 Grumman High School

Grumman High School consists of grades 9-12, and there are about 1500 students in the school. Three elementary schools and two middle schools feed into this one high school that serves the whole town, meaning that students starting at GHS have been in school with some of their peers since early childhood but are joining up with about half of them for the first time. The school rates very well on local and national measures of success. There is a very high (close to 100%) graduation rate, with a similar rate of students going on to attend college.

There are many extra-curricular activities offered at GHS, ranging from sports teams to academic clubs. One which somewhat spans these two types of after-school activities is the Robotics Team. The focus of this club is building a robot to compete at an annual robotics competition. Recently the official name of the group, given to it by the school, has been changed to the "engineering club." This rebranding came as a way to attract more students with a broader term, and also to include a key word that both students and parents are looking for in order to build good college applications. But the club's focus and structure has remained the same.

2.6 The CyberPilots

GHS's robotics club is a member team of the FIRST Robotics League (http://www.usfirst.org/). Called the CyberPilots, students meet after school in the school facilities at least once a week throughout the year, and often every day during the six week long "build season." Each year, on the first weekend of January, the national FIRST organization announces the game for that year –that is, the specifications for how the robots must be built and how they will have to perform for that year's competition. The students then have the build season for designing, building, and testing their robots before the regional competitions begin. They do all this under the supervision and guidance of teachers employed by the schools as well as other mentors. "Mentor" is a defined role in FIRST robotics, as well as an accurate descriptor of the actual job. Typically, mentors are teachers, parents, and other members of the community with special expertise, who volunteer their time to contribute to the team.

2.6.1 FIRST Robotics

FIRST (For Inspiration and Recognition of Science and Technology) was founded in 1989 by Dean Kamen (famously, the inventor of the Segway scooter). The first competition took place in 1992, with 28 high school teams competing their robots in a high school gymnasium in New Hampshire. As of the present day, there are over 3,500 teams registered in 28 countries, and FIRST is backed by many large international engineering corporations who give college scholarships to students. The organization has also expanded to include programs for elementary

and middle school children, in the First Lego League (FLL) and Jr. FLL, where children learn basic building and programming using Lego Mindstorms.

FIRST was founded to promote high school students' interest in science and engineering. It has a distinctive set of guiding principles. The organization is structured around inter-school competitions, yet there is an emphasis on cooperation; the term "coopertition" is used to describe the nature of these inter-school meets. The competition is set up so that groups from different schools are teamed up at the competition and have to work together. FIRST has a commitment to good sportsmanship and "gracious professionalism," and these values are continually reinforced. It is not uncommon to see teams at a competition lending equipment to other teams who they are ostensibly competing against or helping other teams repair malfunctions to their robot.

FIRST is also strongly oriented towards promoting STEM research and education for children of all ages. Teams are expected to perform at high levels not only in engineering skills, but also in doing community outreach and fundraising to promote FIRST's guiding principles and STEM in general. Several awards are given at competitions that go beyond the game. The most prestigious of all awards given at competitions is the Chairman's Award, which earns the winning team an automatic spot at the national competition, even if their robot didn't win the game. The Chairman's Award is judged based on all the outreach and STEM education that the team has performed during the previous five years. The fact that this award occupies the top spot highlights the importance that FIRST puts on valuing this aspect of the organization as equal to, if not more important than, the acquisition of engineering skills.

2.6.2 Prestige and Gender

The population of students involved in FIRST skews male, as many science-and-technology-oriented activities do. In the CyberPilots, all of the teachers and mentors are male and about two-thirds of the students are male. There are still a significant number of female students involved, but they are not distributed evenly among the different roles in the organization. Despite FIRST's stated mission to advance community outreach, prestige lies squarely with the builders and not with the students working on community outreach and fundraisers. In the CyberPilots, the make-up of these two arms of the club strongly correlate with gender – boys building the robot and girls tending to do more of the organization of community programs and fundraising efforts. Officers are chosen each year by the teachers and are announced at the end-of-year barbeque. The student heads of each engineering department (electrical, mechanical, etc.) are largely seen by both students and teachers as the "rightful" pathway to becoming president or other leadership roles. In my observations, this convention does not primarily take into consideration the actual leadership skills that any particular student possesses, and because of the gender skew, also tends to systemically discourage girls from applying for and getting leadership positions.

Chapter 3: Methodology

In this chapter, I describe the approaches I used to gain access to the community and to collect data. I also examine my personal relationship to my research subjects and how that might affect the data I obtained. Finally, I explain how I selected the two linguistic variables—the short-*a* split and THOUGHT-raising—for analysis. The detailed methods used for measuring and analyzing those variables will be discussed in the chapters devoted to each variable.

3.1 Entering the Community

My father has been involved in FIRST Robotics as a mentor since 2003, the year that my brother joined as a student, and he is well-known within Long Island's FIRST community. My brother has also continued to be involved as a mentor ever since he graduated. My father was my entrée into the community, allowing me to start with an advantage that I would not have had if I had been a complete outsider trying to access the group. My family's involvement in the FIRST community is one reason that I chose it as the site of my fieldwork. I have become very familiar with the organization and the community, and I knew it offered the researcher a group of high school students who spend a lot of time together on a regular basis. It also has a somewhat unique trait for a school group, in that it welcomes outside community members to become involved, so coming in as a researcher would not be completely out of place.

At the Long Island regional robotics competition in late March 2015, my father took me around and introduced me to teachers and mentors from various high schools. I was introduced as his daughter who was about to start working on a dissertation and who would like to involve the FIRST community. I told the teachers and mentors who I met that I would be doing a study on high school students on Long Island, how they communicate, and how they use social media and electronic communication. Everyone was very friendly, and several teachers expressed interest and gave me their contact information so that I could follow up with them. I attribute part of the teachers' ready willingness to participate in the study to the FIRST organization's firm commitment to volunteerism and supporting the advancement of science.

I originally hoped to study students at several high schools in Nassau County, but I encountered insurmountable red tape in getting approval from school districts to work with students. The robotics community embraced me immediately, but the school administrations did not feel the same way. I only had success in one district, where I and my family are known. Thus, I adjusted my focus from comparing a handful of participants in several schools, to studying in-depth a larger number of students that comprise one team. This had both negative and positive effects on the potential work. On one hand, the subjects are less demographically varied than if I could have sampled several different school districts, and I cannot compare different areas of Nassau County that are geographically closer or farther away from NYC. On the other hand, all of my subjects come from the same speech community, and indeed the same Community of Practice. I have much more detailed data to analyze this community thoroughly than if my resources had been spread among several schools. I was able to devote more time to

forging bonds with my participants and becoming a contributing member of the team. Plus, I did not have to worry about who to root for at the Long Island regional competition.

In the fall of 2015, I joined the CyberPilots as a mentor. I was officially registered with the FIRST organization as a mentor of the team. At first, I felt like I did not deserve this title, and that I was just pretending to be a mentor so I could collect data. I do not have a background in engineering (although I come from a family where it has always been a part of my life. My father's response to any question I asked as a child was "let's do an experiment!"), but I came to learn that there is much that these students could benefit from guidance on. The scientific method itself, for example, is something they are still learning, and a subject in which I was able to offer expertise. I was also an example to them of someone actually doing science, even if it didn't resemble exactly what they have come to know as "science." Hopefully I opened their eyes to social science as a discipline.

I believe that the most significant impact that I had on the team during my time doing field work was recognizing the leadership skills of those working on the outreach events, who might not have been as hands-on in building the robot, and supporting these students, as well as bringing them to the attention of the mentors as important and capable members of the team. The teachers and most mentors of the club are recruited for their expertise in engineering, so there are often not enough resources to allocate towards overseeing the other aspects of running a club, applying for grants and awards, and aspects of interpersonal organization that adolescents learn from being part of a team. These were some of the roles I was able to help fill during my time as a mentor. It is important for sociolinguistic researchers to not only take from their subjects of

study while doing field work, but also make sure that they are giving something back. I am happy to be able to look back on my time with the team; I feel that I did make some impact.

3.2 Participant Observation and Status of the Researcher

A main concern in sociolinguistic work is always the issue of the Observer's Paradox (Labov 1972), which refers to the problem that, while the researcher's aim is to observe the most natural speech of the speaker, the mere act of observing them has an impact on how they speak and can affect the results. Many aspects of sociolinguistic methodology have been developed with the purpose of minimizing the problem of the Observer's Paradox, most notably the careful design of the sociolinguistic interview to elicit speech that is as unmonitored as can be reasonably achieved. It is also for this reason that many researchers choose to become participant observers in the community of study, to gain not just a more complete knowledge of the social interactions among the subjects to enhance the analysis, but also to become familiar to the community and gain the trust of its members so that later discussions and interviews can be more natural with less attention paid to speech.

I spent the school year of 2015-2016 attending meetings of the robotics club after school and in the evenings. These meetings started out sparsely in the fall but ramped up as build season approached. During the six weeks allotted for designing and building the robot for competition, some number of students could be found in the machine shop every day, including Saturdays. I was not able to attend every single day, but I maintained a presence at least once or twice a week. During this time, I pitched in on small construction tasks where I could be useful, and also

worked with the students who were writing essays and applications for the various presentations and awards that go along with the team participating at the upcoming regional competition. While doing these things, I talked with the students and casually collected information on what the dynamics of the club were like, what their goals were for being involved with robotics, and what their concerns are. This experience greatly informed the design of the sociolinguistic interviews that I conducted in June 2016 and allowed me to tailor my interview modules to topics that were salient to these particular students and would elicit personal and engaging anecdotes. I also built a rapport with the students and mentors that allowed me to overcome the barrier of getting fourteen-year-olds to get a permission slip signed and actually bring it back, as well as schedule a meeting with me and remember to show up. These turned out not to be easy feats.

It is also important for the researcher to consider how their own identity relates to those of the community of study and reflect on how they are seen by the participants, and what effects this might have on the nature of linguistic data obtained. For this reason, I give a brief biography of myself here, and how my own linguistic and demographic history relates to those of my community of study.

I grew up in Nassau County and am an alumnus of Grumman High School myself. My family was also middle-class, white with some Jewish heritage (my mother is Catholic and my father is Jewish, so I had one foot in each identity), and so fairly typical of the area and very similar to the demographics of the speakers in my study. My grandparents all grew up in New York City and moved to Long Island after World War II, when there was a growth of the suburbs. My parents were both born while their families lived in New York City, but moved to

Long Island when they were a few years old – my father to Nassau County and my mother to Western Suffolk County on the south shore. They met each other at a New York State college that had a high population of students from Long Island. After college, they moved back to Long Island to be near their families while growing their own family. I grew up with almost all of my family very close, with my many cousins all on Long Island and with some older relatives that we would regularly visit in Brooklyn and Manhattan. This is a familiar pattern for many of the families in Antioch, and in Nassau County more widely, and in this way, I share a general foundation of life experience with the students I interviewed that allowed me to relate to them about growing up and attending high school on Long Island.

Contrastingly, I am no longer a teenager, and this fact became extremely obvious the minute I started spending time with them. The extent to which communication in the daily life of teenagers has changed since I attended high school was greater than I had previously realized. During my high school experience, cell phones were just starting to become prevalent (I got my first Nokia brick phone in eleventh grade and did not have the ability to text message until after college). Needless to say, there was no social media, and certainly not any that was easily accessible throughout the day on a mobile device. The ways in which the students in this study communicate and connect with each other on a regular basis are influenced by these current technologies, and so constitute a very different framework from when I was in high school. Comfortingly, much is still the same, though. The students in the robotics club still meet after school every day and work together. They go to competitions together and sit in the stands bonding. I found that I was able to relate to them a fair amount, but due to my advanced age, I remained enough of an outsider to their daily technological and communicative practices that I

could believably ask the questions that I wanted to pose as a researcher. (For example, I obtained a narrative from nearly every student about how the social media app Snapchat works. The first one or two times I asked it as a genuine question, but I kept asking it as a way to elicit a similar narrative on a similar topic from each speaker, so now I am quite an expert.) My presence in the robotics shop was generally accepted, as there are always mentors that are not teachers there, although I was the only female mentor during the time I was there. There is regular involvement by parents as well, predominantly mothers, but this was another category I did not quite fit neatly into.

3.3 Interviews and Informants

My data comes from sociolinguistic interviews conducted with 24 student participants and seven mentor participants during the 2015-2016 school year. My time observing the group began in mid-fall of 2015 as the club was organizing and gearing up for kickoff, which occurred on January 9, 2016. I was with the club through the build season and competitions that took place in March and April. My interviews were conducted mostly during June, after the competition season was over and the school year was winding down.

Each sociolinguistic interview was about one hour long, which includes around 45 minutes of conversation followed by a Reading Passage and a Word List task. The interviews focused on topics related to the robotics club, but also school life and family life. I saved language-related questions for the end (Labov 1972), asking individuals whether they thought they had a New York accent, if anyone had ever told them they sound like they are from New

York, and if they thought Long Island had its own accent or not. A sample interview module is included in Appendix A.

As a reading passage, I chose a story told on a talk show by comedian Louis C.K. (this was before the #MeToo movement happened), and then added additional tokens of THOUGHT and short-*a*. I wanted something that sounded natural and was familiar in content and style to the students, rather than a constructed passage that was distractingly contrived. The word list was 147 words long and adapted from the word list in Newlin-Łukowicz 2015. Word order was randomized, and each speaker received a sheet of paper with the words in a different order. I made sure there were not any obviously related words next to each other that would have induced a conscious comparison between the words by the speaker, as in a minimal-pair reading task. The reading passage and word list can also be found in Appendices B and C.

The interviews took place in various rooms at the high school and also at the Antioch Public Library, where it was possible to reserve a private study room. At the high school, some interviews were conducted in a small, quiet room adjacent to the robotics machine shop. When the shop was not open, I reserved a room elsewhere in the school — sometimes a back room in the school's library (which also happened to be a Holocaust research center), and sometimes a conference room in the administrative section of the school. I interviewed two of the mentors who were not teachers in my parents' living room, as it was easier for them to come over after work when the season was over, and they were not at the school all the time.

Additionally, some interviews were conducted while at the regional competition in the Hofstra University field house. I was able to obtain access to a side room outside of the gymnasium. This was a fruitful time to schedule interviews because all of the students were

there, and it was easy to arrange several interviews on a given day without inconveniencing the students too much. These interviews have some noise interference, however, because despite signs on the door that said the side room was off-limits, people still intermittently opened the door to look in and realize that it was not where they were supposed to be, letting in the roar of the crowd from the gym. Using a directional lavalier microphone, I was still able to capture audio that was high quality and adequate for phonetic analysis. Interviews were collected with a Zoom H4n portable recorder and an Audio Technica lavalier microphone. All interviews were recorded in WAV format with a sample rate of 44.1 kHz.

3.3.1 Students

The 24 student participants in this study are listed in Table 1 below. They are organized by age from youngest (Freshman) to oldest (Senior). Within each grade level, the girls are listed first, followed by the boys.

Name	Age	Gender	Ethnicity	Religion
Diviya	Freshman	Female	Indian	Sikh
Rebecca	Freshman	Female	White	Jewish
Chloe	Freshman	Female	White	Jewish
Ranjit	Freshman	Male	Indian	Hindu
Tim	Freshman	Male	White	Jewish
Alex	Freshman	Male	White	Jewish
Caleb	Freshman	Male	White	Jewish
Tahani	Sophomore	Female	Indian	Sikh
Ethan	Sophomore	Male	White	Jewish
Jake	Sophomore	Male	White	Jewish
Noah	Sophomore	Male	White	Jewish
Carina	Junior	Female	Chinese	
Hannah	Junior	Female	White	Jewish
Leah	Junior	Female	White	Jewish
Luke	Junior	Male	White	Jewish
Joe	Junior	Male	White	Jewish
Adam	Junior	Male	White	Jewish
Logan	Junior	Male	White	Jewish
Amy	Senior	Female	Chinese	
Christina	Senior	Female	White	Catholic
Matan	Senior	Male	White	Jewish
Garrett	Senior	Male	White	Jewish
Daniel	Senior	Male	White	Jewish
Paul	Senior	Male	White	Jewish

Table 1: Demographics of Student Participants

Among the students there are seven freshman, four sophomores, seven juniors, and six seniors. The girls make up 38% of the sample (9 students), while boys make up 62% (15 students). Three-quarters of the students are white (with eighteen of those being Jewish and one being Catholic). I include religion in this table because in most cases it is a salient component of the students' ethnicity. The majority of Antioch, and of my participants, are Jewish. Labeling them simply as white does not fully describe their ethnic identity and would not capture the

difference in identity between Jews and Catholics (who are usually Italian) on Long Island. Similarly, among the three Indian students in the study, two are Sikh and one is Hindu, which is a salient difference to them, and influences who they spend time with.

I want to draw attention to Matan here. I have categorized him as white for the purposes of this study, but he does differ from the other students in being Sephardic while all the other Jewish students are Ashkenazi, the sect of Judaism that flourished in Eastern Europe and the historically more common one to have been brought to the US by immigrants. Sephardic and Ashkenazi traditions have been separate for centuries, and there are different cultural and religious practices associated with each. It would be an incomplete description of Matan to not mention this, as both his name and his visual appearance mark him as middle eastern, and his upbringing has been different than those of his peers because of it. Despite this difference in culture, I did not find Matan to behave linguistically differently from the other Jewish students, so I included him in the same group in the analyses that follow.

I have not listed a religion for the two Chinese students, Carina and Amy. This is because it never came up in conversation with them, as all of their answers to questions about family, holidays, and traditions were about their Chinese heritage, and any religious beliefs held by them or their parents were not relevant enough to them to mention. In her interview, Carina told me specifically that she and her family identify as Chinese first and foremost, even though her father is white, and she refers to herself as “a hybrid.” When asked about holidays, Amy said that her family celebrates Christmas, but not in a religious way at all, but purely for cultural reasons, and that the most important holidays to her and her family are Chinese.

3.3.2 Mentors

The mentor group consists of seven participants who are adults who help run the robotics club. Three of these mentors are teachers who are employed by the school and who are paid for their responsibility as club advisors. These three teachers are referred to as "Mr. Lastname" in order to differentiate them from non-teacher mentors, who are referred to by their first names. The mentors are listed in Table 2 below in age order.

Name	Age	Gender	Ethnicity	Religion
Anthony	24	Male	White	Catholic
Mr. Price	30	Male	White	Catholic
Nate	31	Male	White	Jewish
Isaac	59	Male	White	Jewish
Mr. Spector	60	Male	White	Jewish
Mr. Geissler	61	Male	White	Catholic
Lenny	86	Male	White	Jewish

Table 2: Demographics of Mentor Participants

The mentors are a more homogeneous group than the students. They are all male and all white. Three are Catholic and four are Jewish. They cover a large age range, from 24 to 86 years old, but there are three main age clusters. Tony, Mr. Price, and Nate are the younger mentors, in their 20s and early 30s; Mr. Spector, Mr. Geissler, and Isaac are the middle-aged mentors at around 60 years old, and Lenny is the oldest mentor at 86 years old at the time of the interview.

Of the mentors, only Tony was born and raised in Antioch. Mr. Price is from a nearby town on the south shore and Nate is from a town in western Nassau County. Isaac was born in the city and moved to Nassau County when he was about two years old. Mr. Geissler is unique in that he is the only participant, mentor or student, whose family has been on Long Island for

several generations before WWII, and before the suburbanization in the area began. His Long Island roots are a bit deeper than anyone else's in the study.

Both Mr. Spector and Lenny did not grow up on Long Island. They were born and grew up in Brooklyn, moving to Long Island in their early twenties (25 years apart). I include them in this study because their language is part of what children like the CyberPilots hear growing up in Nassau County, not just when they go to visit their relatives in Queens, but in Nassau itself. Many of the students have parents who grew up in Brooklyn and Queens, and it is a normal part of their adult language input from parents and teachers alike. I do not include Mr. Spector and Lenny in statistical analyses that compare mentors to students, limiting those comparisons only to Long Island born speakers. Elsewhere I make it clear when it is relevant that Mr. Spector and Lenny grew up in Brooklyn.

As described above when I told the history of my own family, it is a common pattern for people who live in Brooklyn and Queens to move to Long Island once they have one or more children. This is true of myself; I lived in an apartment in Bayside, Queens with my parents until I was about six months old. Needless to say, I do not remember this time, and I consider myself to have been born and raised on Long Island. This is true of some of the students too. Below in Table 3, I list the students who mentioned in their interview that they had been born somewhere other than Antioch and moved there as a small child. However, given the prevalence of this pattern, I am not certain that there are not other students who would fit into this group that did not mention it when I asked them where they were born, as I myself would not answer Bayside to this question.

Student	Born In	Age Moved to Antioch	Year in School	Gender	Ethnicity
Matan	Queens	1 year	Senior	Male	White Jewish
Garrett	Brooklyn	1 year	Senior	Male	White Jewish
Ranjit	Queens	1 year	Freshman	Male	Indian
Diviya	Queens	2 years	Freshman	Female	Indian
Amy	Queens	2 years	Senior	Female	Chinese
Noah	Queens	5 years	Sophomore	Male	White Jewish
Chloe	Queens	6 years	Freshman	Female	White Jewish
Tahani	Queens	<10 years	Sophomore	Female	Indian

Table 3: Students who were not born in Antioch

Of the student participants, seven reported being born in Queens and one in Brooklyn. The table above lists the ages at which they moved to Antioch - all are well within the window of language acquisition. I do not have an exact age at which Tahani moved to Antioch, but I know that it was before she was ten years old. She had also spent lots of time in Nassau County before that, at the large Sikh temple in the area. Tahani actually spoke in her interview specifically about feeling that NYC and Long Island were different and that she definitely saw herself as a Long Islander.

3.4 Linguistic Variables

The phonetic variables pertaining to New York City are well studied. The five that Labov (1966) investigated are: *r*-vocalization, THOUGHT-raising, the short-*a* split, TH-stopping, and DH-stopping. From initial interaction with the high school age participants, I was able to eliminate

TH- and DH-stopping conclusively, as no one in this community of any age seems to employ this variable regularly. I was interested in *r*-vocalization, because my initial observation was that it is produced at moderate to high rates among the generations above my student participants, but hardly at all by the students themselves. Analyzing this variable would thus have allowed me to potentially highlight a change in progress, or perhaps one that has virtually reached completion, wherein the youngest Long Islanders no longer use vocalic /r/s. However, features whose variation is more robust seemed to me to have the potential to reveal more about language use on Long Island, especially with regard to the relationship of New York City English to Long Island, so I set aside *r*-vocalization in order to focus on the two remaining canonical features: THOUGHT-raising and the short-*a* split.

Chapter 4: A "Short" History of "Short-*a*" in New York City

4.1 Background

In American English, the low front vowel /æ/ often surfaces with two allophones, a "*tense*" version (raised and fronted) and a "*lax*" version (lower and backer), as shown in Figure 8 below.

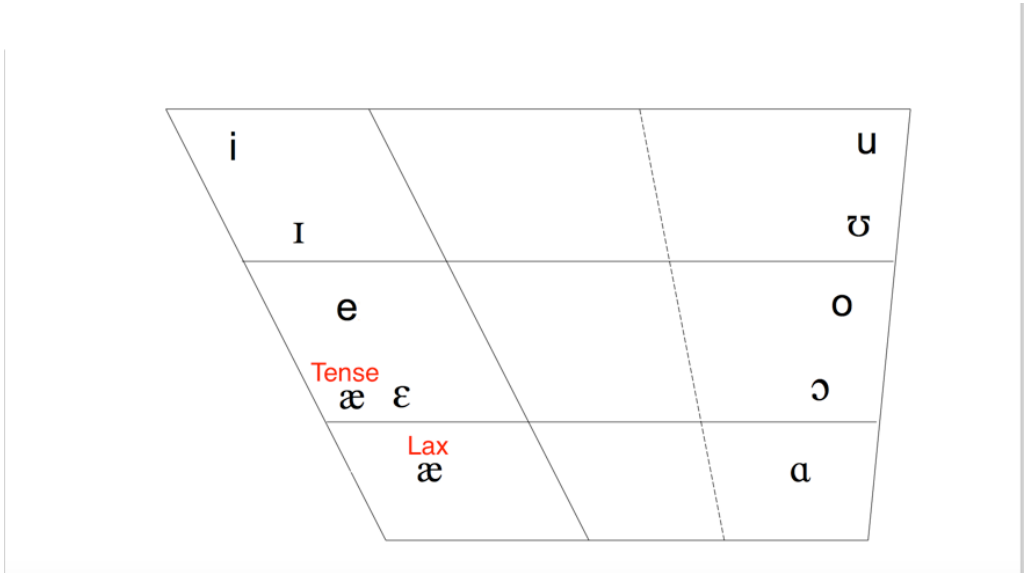


Figure 8: An abstract representation of the English vowel space with the positions of *tense* and *lax* short-*a* included

There are several different systems of phonological conditioning of these allophones that exist in various dialects. The most common pattern in American English for the pronunciation of /æ/ is known as the "nasal split" (ANAE 2006:175). In this system, the vowel is *tense* when followed by a nasal consonant and *lax* otherwise. Figure 9 shows the vowels of a speaker with a nasal-split

configuration. All of the tokens in this graph that include a short-*a* followed by a nasal, either front or velar, are on or above the green horizontal line, which appears to be placed at around 670Hz. Below this line, and also farther back, are tokens that contain any following environment that is non-nasal.

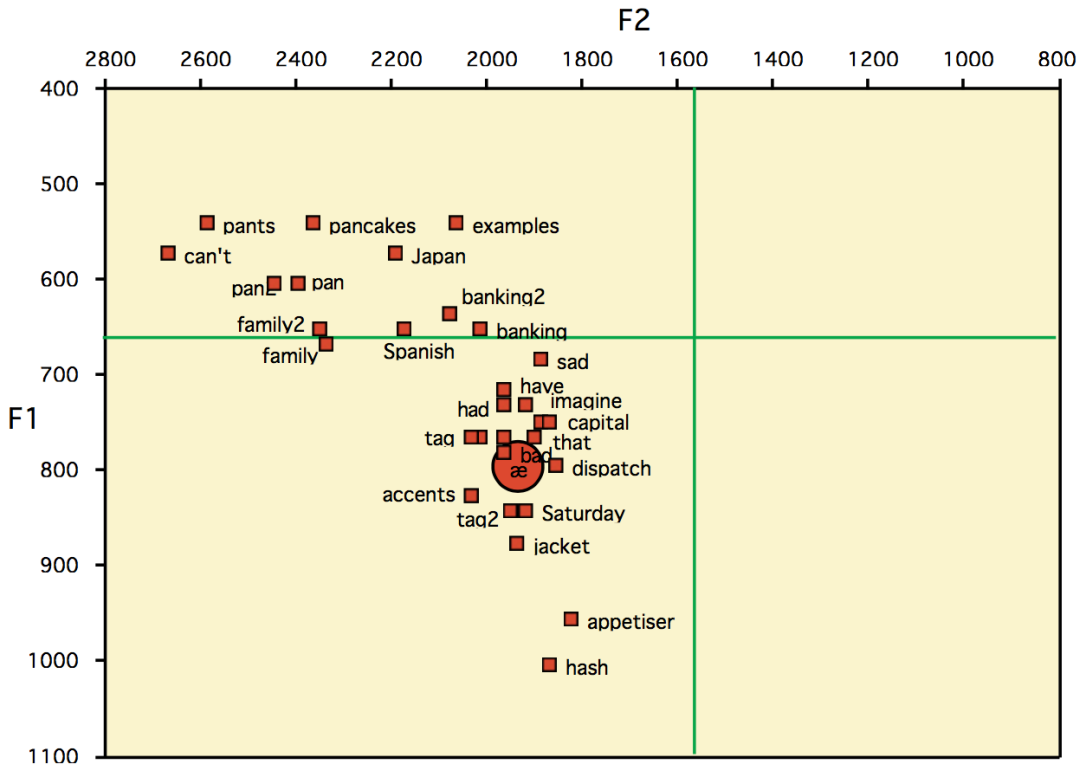


Figure 9: The nasal short-*a* configuration in the vowel system of Danica L., from the Atlas of North American English (2006:175)

New York City English (NYCE) has had a historically "complex split," described in detail below. A complex split still shows patterning based on following phonological environment, but the conditioning environments are more involved than "pre-nasal vs. not." An example of this kind of split system can be seen in Figure 10 below. There are two distinct

clusters of pronunciation of short-*a* represented on this chart, one above the green line and one below.[4] However, unlike the plot below showing a nasal-split system, the words located above the line include additional following environments such as voiced stops ("bad") and voiceless fricatives ("grass"). Additionally, words with velar nasals are actually seen below the line among the *lax* pronunciations (see "Sanka" and "Frank").

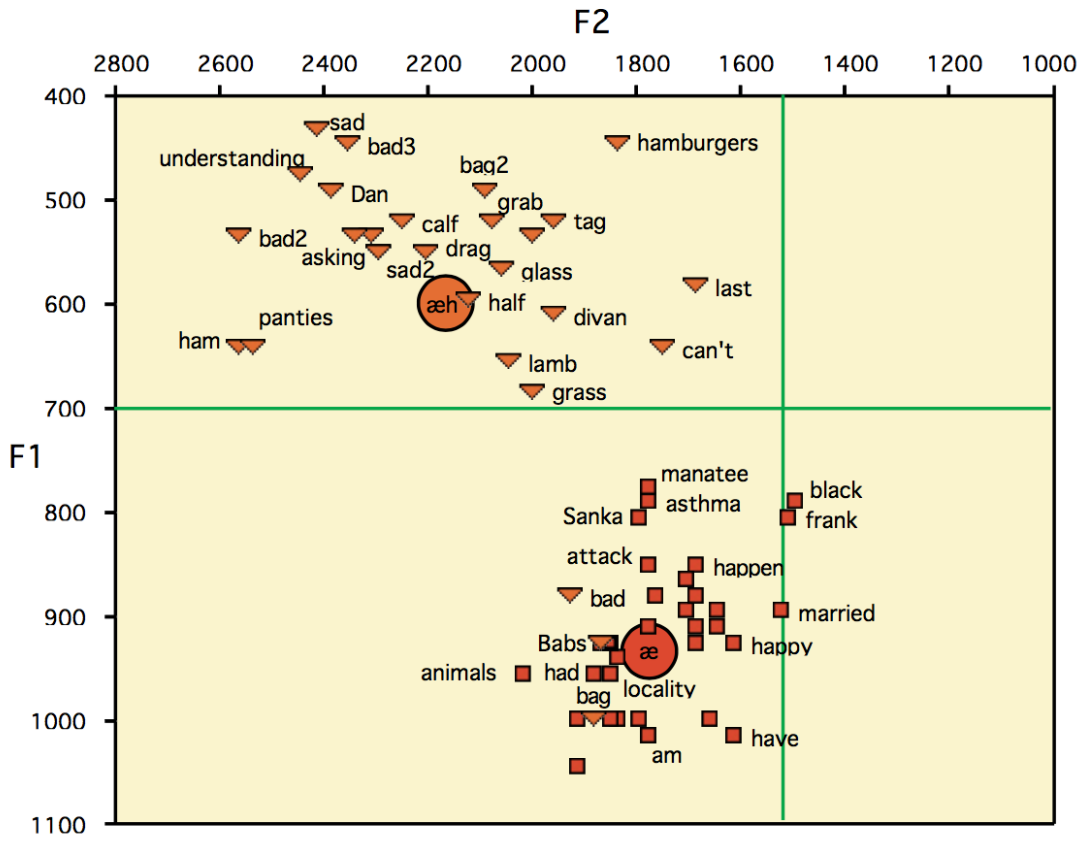


Figure 10: Split /æ/-/æh/ system for Nina B., from the Atlas of North American English (2006:174)

[4]Note that the benchmark for determining if a short-*a* is raised is given at 700Hz in the Atlas of North American English, but it varies from graph to graph within the book where this dividing line is actually drawn. It is always somewhere between 600Hz and 700Hz.

A third possible system that some American English speakers exhibit is a "continuous raising system." In this system, short-*a* before front nasal codas are the *tensest* and short-*a* before voiceless stops are generally the *laxest*, but there are no significant gaps in the distribution of tokens between these two extremes (Labov 2007, Becker 2010). See an example of a continuous raising system in Figure 11. In this figure there is a green line drawn (here around 630Hz), but the tokens do not clearly cluster on either side of it with a gap in between.

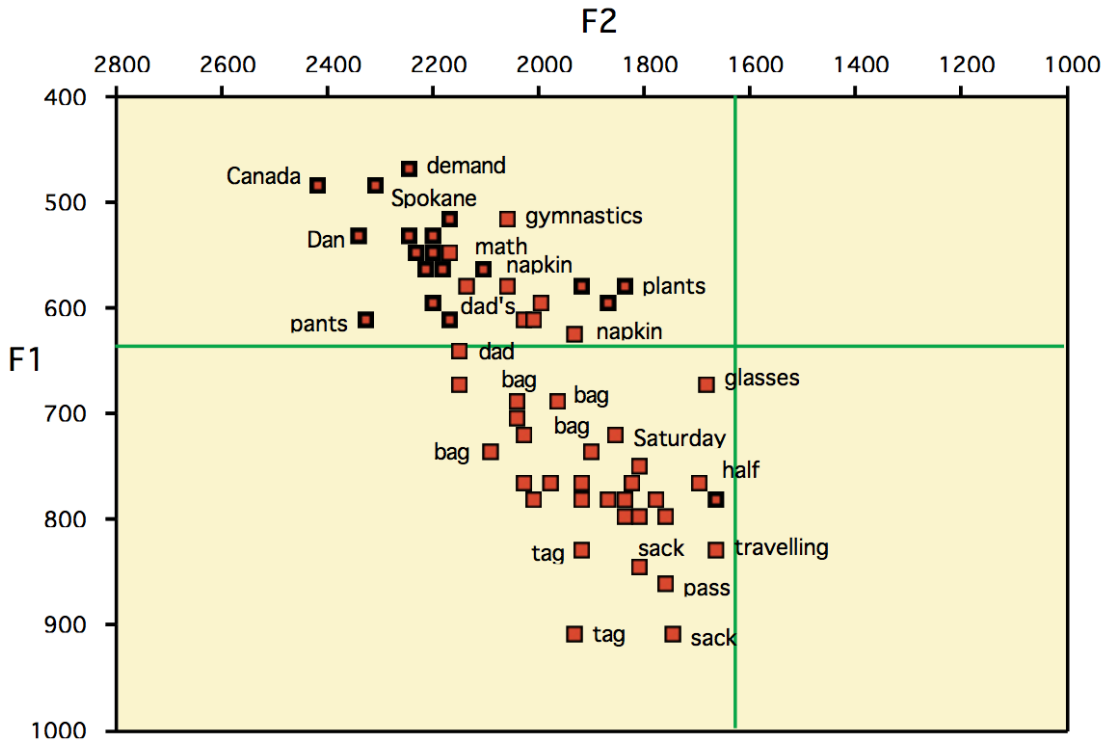


Figure 11: Continuous short-*a* system of Lorrain K., from the Atlas of North American English (2006:179)

4.2 Literature Review

In reviewing the history of the study of the short-*a* vowel in New York City, I start with the early dialectological work, based largely on the authors' native intuitions, that first brought to light the complexity of the split in New York City more than 100 years ago. I then summarize the early variationist work that empirically looked deeper into the complex split and worked to describe it in detail. There was then a gap in focus on this variable in the field for about thirty years after which William Labov and colleagues famously took up the mantle again in *The Atlas of North American English* (2006). Following *The Atlas* and Labov 2007 there was a renewed interest, spurring a series of papers and dissertations examining the state of short-*a* specifically in New York City (many of them out of the NYU linguistics department). Most recently, Elizabeth Coggshall's 2017 dissertation focused entirely on short-*a* in NYC and New Jersey. She reexamined and reframed the entire history of study of the variable, and that is the starting point for my own analysis.

Much of the narrative that I build in this chapter I owe to Coggshall 2017 and her exhaustive review of the literature. The crux of her dissertation is a convincingly built argument that recent work on short-*a* in NYC has not accounted for variation that was always present in the system and had been documented in early work, but that was later glossed over in order to create a more streamlined description of the phonological system. Throughout my own review, I also highlight areas where variation in the production of short-*a* was noted in certain phonological contexts and then point out where later papers, relying on the simplified version, used instances of this same variation in their own data as spurious evidence of the collapse of the

system. In my own data analysis, I attempt to maintain a keen awareness of the variation that Coggshall astutely brings back into the discussion, while still recognizing the need for some simplification in order to make sense of what Coggshall calls "perhaps the most complicated sociophonetic variable in American English" (2017:226).

4.3 Foundational Dialectological Work

4.3.1 Babbitt 1896

The earliest descriptions of the New York City dialect, and specifically the pronunciation of the short-*a* vowel, came from dialectological work. The split of New York City's short-*a* into *tense* and *lax* sets was first described by Babbitt in his 1896 paper "The English of the Lower Classes in New York City and the Vicinity." Babbitt noted the raised-æ (*tense* short-*a*) where "all the set of words pronounced in New England with the broad vowel (ask, half, pass, etc.) occur, and is really higher in these words than in "man, cab" (461). Babbitt does not give much more detail or any specific data (his work is based on "quietly taking notes on [the local pronunciation] during [his] six years of residence in New York" (460)), but he seems to be placing the class of words with following voiceless fricatives as *tenser* than that of voiced stops and front nasal, while all three are in his category of raised short-*a*.

Notice that the title of Babbitt's paper includes "and the vicinity" when talking about the range of the New York City dialect region. He says that "only a historical accident…separates [Manhattan Island] from the part of New Jersey across the river which looks to Manhattan Island

for its business interests" (457). Although Babbitt did not write about Long Island, this description is true of Nassau County as well.

4.3.2 Trager 1930, 1934, 1940

Trager gave a more in-depth analysis of the short-*a* split in New York City in a series of papers in the journal *American Speech* between 1930 and 1940. Using his own intuitions as a native speaker from the region (he's actually from Newark, New Jersey, not New York City proper), he gives the first thorough description of the following environments and additional factors that condition the two variants of the vowel in the split. In recent work, the description of the NYCE short-*a* split has become so strongly tied to the work of William Labov that most current work on the variable refer to it as the "Labovian" split (e.g. Becker and Wong 2010, Becker 2010). As Coggshall 2017 establishes, however, Trager's work was the original foundation of the descriptions that Labov drew on in his work from 1972 onward and, in fact, that researchers are still drawing upon today. Therefore, I follow Coggshall in referring to the complex short-*a* split in NYC as the "Tragerian" split (116). As Coggshall says, "Trager is the benchmark of authenticity" (116) for much of the work done on the short-*a* split in New York City since 1940.

In Table B on page 398 of his 1930 paper (reproduced below as Figure 12), Trager lays out all of the relevant following phonological environments and whether he himself, a self-described speaker of "Standard American English" (1930:396), has *tense* or *lax* short-*a* before each of them.

TABLE B

1. æ	always before p, t, k, tʃ, ŋ, l, r.
2. æ	before b, d, g, f, ɵ, s, z, when these are followed by consonant plus vowel, or by vowel alone (except in inflectional forms).
3. æ	before dʒ, m, n, v, ð, when these are followed by a vowel (except in inflectional forms).
4. æ	in the words *had, have, has.*
5. Æ:	always before ʃ.
6. Æ:	before b, d, g, dʒ, m, n, f, v, ɵ, ð, s, z, when these are final or followed by another consonant which is final; or when followed by a vowel of an inflectional ending.

Figure 12: Table of following phonological environments conditioning either *tense* or *lax* short-*a*, from Trager (1930)

Trager describes short-*a* as always being *lax* before following /p, t, k, tʃ, l, r, ŋ/, and always being *tense* before /b, d, g, dʒ, m, n, f, v, ɵ, ð, s, z/ in closed syllables. That is, short-*a* is *lax* before voiceless stops, voiceless affricates, liquids, and the velar nasal. It is *tense* before voiced stops, voiced affricates, voiceless fricatives and front nasals. Trager also notes that short-*a* is *lax* when followed by any of the above following environments plus a subsequent vowel. We now describe this as the effect of being in an open syllable, but Trager did not use that terminology. Trager gave two exceptions to the open syllable rule: he singled out intervocalic /ʃ/ and the cases where the subsequent vowels are part of inflectional endings. According to Trager (1930:399), following /ʃ/ sometimes conditions a *tense* short-*a* (e.g. in *fashion),* but sometimes doesn't, as "*passion* seems to have [*lax* short-*a]* very often." Here, Trager is recognizing that there is variation in the production of short-*a* before intervocalic voiceless fricatives. This is also

the first mention of what Labov 2007 later called "inflectional boundary closing," or the phenomenon where the subsequent vowel after the following environment does not cause an open syllable effect if that vowel is part of an *inflectional* morpheme specifically. Trager attributes this to "analogical levelling" (1930:399). In Figure 12, Trager also lists the function words "have, has, and had" as exceptionally having *lax* short-*a* despite following environments that usually condition a *tense* short-*a*.

Trager tries to find a phonetically grounded reason for this complex pattern, and he suggests that following voiced stops and voiceless fricatives might lengthen and *tense* vowels in American English in general (399). Because function words are seldom stressed, this would not usually apply to them, leaving the short-*a lax*. Trager proposes that since the velar nasal and /l/ are both pronounced "at the back of the mouth," the vowel is consequently backed as well, making it *lax* rather than *tense* (399).

In his 1934 paper, "What conditions a phoneme?" Trager uses [ɛ] instead of [æ] and says that there are three variants of the variable instead of the previous two: short, *lax*, and open. This new third variant is described as a lengthened-*lax* version and is added to account for what he observes occurring before /l/ and the velar nasal. Trager more explicitly makes the observation in this paper that both following environment and word structure (open vs closed syllables) affect which short-*a* occurs. He notes "halve" and "have" as a minimal pair, but still considers the variants allophones of one short-*a* phoneme. At this time, Trager doesn't have an explanation for why the function word "have" would be *lax* when the content word "halve" is *tense*, except that there may be a "lexico-morphological criterion" (1934:315) that takes into account the meaning and function of a word.

In Trager's last paper on the subject, "One phonemic entity becomes two: The case of short-*a*," (1940), he changes his previous position and argues that the short-*a* variants are actually two separate phonemes. He goes back to representing the two variants as [æ] and [æˑ] (255), and he abandons the third, lengthened *lax* variant that he proposed in 1934 specifically for the following /l/ and velar nasal environments. He still finds *lax* short-*a* before voiceless stops, voiceless affricates, velar nasals and /l/, and *tense* short-*a* before voiced stops, voiced affricates, front nasals, voiceless fricatives, and /v/ and /z/ (255). When the following consonant is intervocalic, the short-*a* will only be *tense* when the second vowel is the beginning of a "regular formative or paradigmatic suffix." Coggshall interprets this to be a reiteration of "inflectional morphemes" as Trager had called them before but as I discuss below, I think this could also apply to derivational morphemes, and it's not clear that Trager is excluding them in this specific formulation of his description.

Trager mentions the existence of the pattern he is describing specifically in New Jersey, NYC, and Philadelphia, but says it "has been noted in the speech of many in widely scattered parts of the United States, especially in cities" (1940:355). As mentioned earlier, Trager's descriptions in these three papers are the real foundation of most of the conditioning factors we use in analyses of short-*a* today, including following environment, syllable structure, inflectional boundary closing, and function words.

4.4 Early Variationist Work

This section describes work done in the 1960s and 1970s that builds on the earlier dialectological work discussed above, and furthers it by adding empirical, quantitative analysis based on corpora of data collected in the field. The first of these that I discuss is Labov's 1966 work "Social Stratification of English in New York City" (SSENYC), a work seminal to the field of sociolinguistics as a whole, and that includes in large part an analysis of the short-*a* variable. Following that, Labov, Yaeger, and Steiner (1972) (LYS) do further analysis on the data from SSENYC. Included here is also a discussion of the work by Cohen (1970), a student of Labov's, who combines analysis of the SSENYC data with his own additional data collected from each of the five boroughs of NYC, as well as from New Jersey. What unifies these works is their use of corpora of empirically collected data, as opposed to the more casual observations and native intuitions of the early dialectological work. They differ from the later variationist work in that the analyses described in this section use impressionistic coding of short-*a* into *tense* and *lax* variants, and do not yet use the instrumental measurement of vowel formants that later work does.

4.4.1 Labov 1966

In SSENYC, Labov looked at the low vowels (both front and back) and their systems of "raising" in New York City. In this work, he refers to *tense* short-*a* as "raised æh." Labov found æh-raising to be part of a larger system of raising in the NYC dialect. In this study, he analyzes variation in the degree of raising of the vowel, rather than describing the split pattern or the

distinction between *tense* and *lax* variants of the vowel. Labov defines three categories of following environment that condition the production of short-*a* but doesn't mention at this juncture any of the additional constraints that Trager did about syllabic structure and the like, although Labov does include them in later work. The analysis here is only of closed, final syllables in *tensing* environments. Labov found patterns in the degree of raising (i.e. *tensing*) based on ethnicity, social class, speaking style, age, and gender.

The speakers in Labov's study were residents of the Lower East Side of Manhattan from three ethnic groups that he deemed to contain a critical mass of native English speakers: Jewish, Italian, and African American. After identifying all of the short-*a* tokens from each speaker's interview, he coded each impressionistically for its height (Labov does not account for fronting in this study). He assigns each participant an index value based on their pronunciations, where a higher index score corresponds to a lower (or *laxer*) short-*a*. A score of 40 or less means a *tense* production of short-*a* (Labov 1966:35). Each of the three ethnic groups have average index scores of less than 40, indicating that they all *tense* short-*a* in the phonological environments that Labov identified as *tensing* environments. Italians had the lowest overall scores, which means they did the most æh-raising.

Labov found that *tense* short-*a* had social stigma attached to it, which he established by comparing the production of the vowel across social class, and also across different styles of speaking. Both of these social variables are represented in Figure 13 below, a reproduction of Figure 7.2 from Labov 1966:143. He found that his lower-class speakers (classes 0-2) were the most raised, the working-class speakers (classes 3-5) were in the middle, and the middle-class (classes 6-9), his highest class of speakers, produced the least æh-raising. Additionally, Labov

found a reduction in æh-raising in more formal speaking styles. Here, Labov defines "style" as "attention paid to speech." His categories are, from least to most formal: speech outside the interview or in emotional moments of the interview (A), most of the speech in the interview (B), reading passages (C), and word lists including minimal pairs (D).

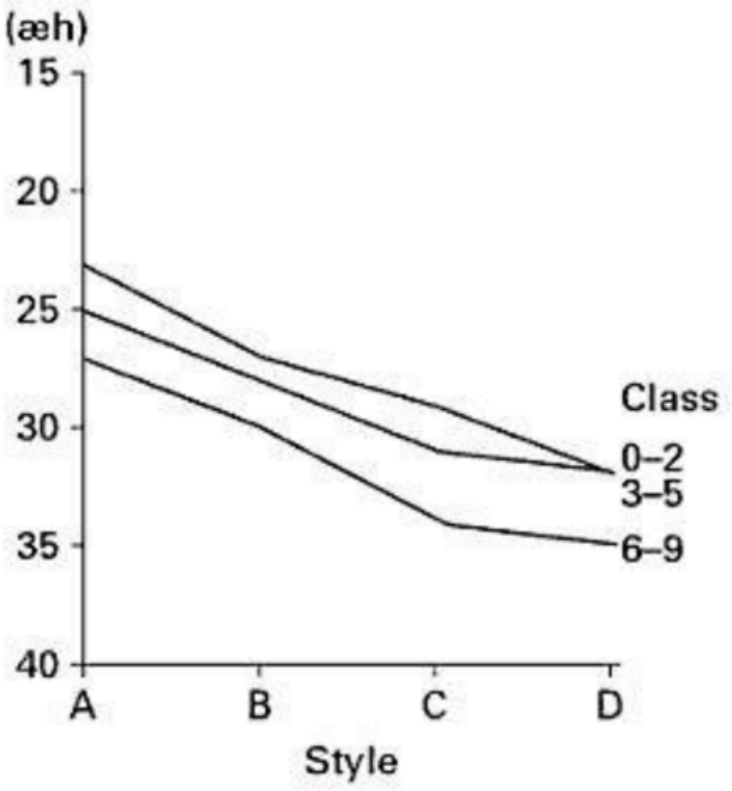


Figure 13: æh-*tensing* index score by economic class and speaking style, from Labov (1966)

Further evidence for the social evaluation of this variable comes from another graph presented by Labov (Figure 14) that is similar to the one above, except for that it separates out the highest social class (Class 9) from the rest.

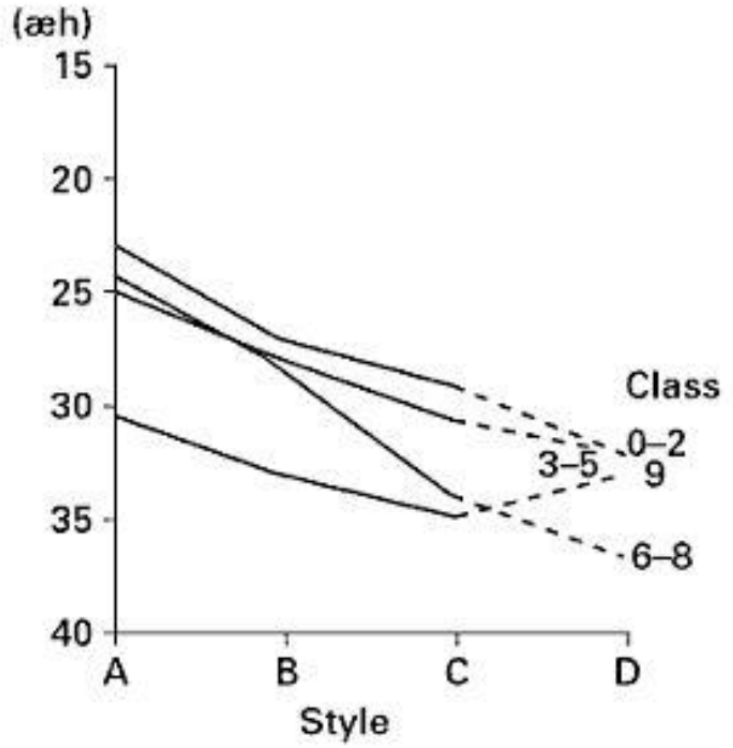


Figure 14: (aeh) index by class and style showing the lower middle class cross-over, from Labov (1966:163)

In this graph, there is a second-highest cross-over effect, with the lower middle class exceeding the level of "correction" to lower short-*a* than the highest class. The cross-over effect describes a situation in which members of the second-highest class avoid the stigmatized variant in order to sound more upper class, while members of the truly higher class do not feel they have to prove themselves.

Labov did not find any style shifting among the African American speakers in his study, suggesting that they might not have felt the same stigma around the *tense*/raised variant as did the Jewish and Italian speakers. While Italian speakers had overall higher pronunciations of

short-*a* than the Jewish speakers, both of these groups reacted the same way to the social pressure of stigmatization and ended up converging on the graph in the more formal styles. Labov concludes from this that at least for white speakers: "New York forms a single speech community, quite diverse in everyday speech, but united by a common norm – in this case, expressed by the standardization of the most formal pronunciation" (SSENYC 186).

Despite the reduced rate of æh-raising in NYC in more formal styles, Labov proposes that it is "a prestige marker with linguistic change in progress" (226). Because the lower classes also participate in this change, which isn't typical for prestige-motivated language change, Labov further proposes that this is a "change from below with the addition of a later reaction from above" (226). The lower class is affected because it is initially a change that is below the level of consciousness, but then the effects of stigmatization are added on top of that, creating the various patterns found with regards to social class and style that Labov observed. With regard to age-grading, Labov found that younger speakers showed a more negative response to raised-aeh (297-298), and also that women he interviewed tended to have higher raised-aeh. Both of these facts fit with the theory that æh-raising is primarily a change from below.

Labov further proposes substratum effects from the Italian and Jewish communities as strengthening forces of this change already-in-progress (192-193). All of Labov's speakers were monolingual and native to New York, but he suggests that the second-generation Italian and Jewish speakers had a desire to distance themselves from their immigrant parents, and so hypercorrected away from the languages of their respective communities. For Italians, this meant æh-raising to move away from the low and back /æ/ of recent Italian immigrants, while Yiddish

has a high and front /æ/ in the language, so their children would move in the opposite direction to establish themselves as different from their "foreign-sounding compatriots" (193).

Labov (1966:227) also proposes a timeline for this change in progress. Since the oldest of his lower class speakers (born around the turn of the twentieth century) do not exhibit æh-raising even in the expected *tensing* environments, he concludes that the current æh-raising trend that he observed began sometime before World War I, and that it first affected Italians in New York City, then Jews, and then African Americans, which is why the stratification between those three groups existed at the time Labov recorded his data (227-228). He found no significant apparent time difference for the Italian group of speakers, leading him to believe that the change had already concluded for this group by that time. Due to Babbitt's description of the vowel in 1896, Labov also concluded that an earlier raising event occurred in the nineteenth century, separate from the change in progress he recorded in SSENYC (227-228).

4.4.2 Cohen 1970

Cohen was a master's student of Labov at Columbia University. He was a native of Washington Heights in northern Manhattan. In comparing his own short-*a* split intuitions to those of Trager, he found very few differences despite the two being born 35 years apart. Cohen isn't often cited, but Coggshall (2017) brings his work into the foreground, calling it "one of the foundational studies on the NYCE short-*a* split" (163). She focuses his work at the center of her discussion because it is the first in-depth study that used empirical data to look at which word classes are categorized as *tense* and which as *lax* by speakers, rather than comparing the relative height of only *tense* tokens, like Labov did in SSENYC. Cohen also looked across a wider swath

of the NYC dialect region, including New Jersey (which is Coggshall's focus), not just limiting himself to the confines of New York City proper (which has really meant "Manhattan" in previous studies). Cohen's work is also of particular relevance to mine because he looked at children and young adults as his subjects, suspecting a change in progress and identifying this age group as a fruitful site of study to document such a change (16). All of his participants were male, under 30 (born between 1940-1955), white, working class, and hadn't spent significant time away from NYC. He wanted to minimize the effect of the social correction that SSENYC had found, so he limited his subjects to the working class, whom Labov had found to be the least affected by the stigma of aeh-raising.

Cohen defines *tense* short-*a* as "higher, longer, *tenser*, and sometimes nasalized, and often an ingliding diphthong" (3). He breaks down his tokens into an impressive twenty-five separate classes of words, based on following environment, syllable structure, and also syntactic boundaries (37-39). He uses impressionistic coding of the vowels to categorize them as *tense* (which he represents as [ɛ:]) or *lax* (which he represents as [æ]).

Cohen hypothesizes that when his speakers use *tense* short-*a* where he expected them to use *lax*, that this is an indication of an ongoing simplification process to the complex Tragerian system. He attributes the additional "complications and anomalies" (67) in the pattern to being a result of this phonological simplification process. Cohen does an extremely thorough job of detailing these patterns.

Cohen concludes that "those farther from New York City have short-*a* splits that least resemble that of the classic Tragerian NYCE split" (Coggshall 2017:164), but his reference for "farther from New York City" is on the New Jersey side, which has a different history and

demographic background than my site of study east of NYC. So it remains to be seen if Cohen's generalization about "farther from New York City" can be extended to the east as well as to the west.

4.4.3 Labov, Yaeger, and Steiner 1972

Labov, Yaeger, and Steiner 1972 (LYS) is based on the interview data from SSENYC. While they also situate the short-*a* split in New York City as part of a larger pattern in the vowel system of English in the US as a whole, they do a much more in-depth analysis of the conditioning environments of the split system itself, and not just the raising trend that Labov 1966 focused on.

LYS presents the basic *tensing* rule as such: "In New York, *tensing* affects /æ/ before front nasals and all voiceless fricatives and voiced stops, if the next segment is an obstruent of [sic] + [morpheme] or # [word] boundary and if the word is not [+weak] [...] i.e., a function word with only one vowel which can be reduced to schwa" (1972: 48). By formulating the rule as such, they recognize the important factor of following phonological environment, while also collecting the first methodical list of exceptions, or additional conditions, that apply to the *tensing* pattern. In fact, LYS lays out a list of 11 exceptions, attempting to account for all the variation present in the system. Their list was the most detailed of any study of New York City English to date (Coggshall 2017:141).

Important to note here are the environments that LYS highlights as variable within the system, because later work makes these patterns seem much more exceptionless than they are presented here and effectively erases the variation described here. LYS mentions specifically that

short-*a* before voiced fricatives such as in "jazz" and "razz" is variably *tense* for different speakers and unpredictable as to which speakers will use which variant (49). Similarly, according to LYS, "a few speakers raise velar nasals, or at least raise them partially, so it is not clear what the class of bang is" (49). By this, they mean that short-*a* occurring before a velar nasal, such as in the word "bang," is not consistently in the *tense* or the *lax* class, according to their analysis. Later work takes raising before velar nasals as one of the primary indicators of the disappearance of the Tragerian short-*a* split, which Coggshall (2017) points out is problematic given the variation in this environment noted in these earlier papers.

LYS both confirm some trends first noted in Labov 1966 as well as propose some new conclusions. As for the former, they confirm that the Italians and the women in the data are the most advanced in the raising of short-*a* (53-54), that younger speakers don't raise as much (56-58), and that there indeed exists a social stigma against the raised variant of the variable. Further, they find that there is a regular progression through the generations of short-*a* variants being more of a continuous range for the oldest speakers, with a slightly higher allophonic variant of /æ/, but then a distinct gap between the two allophones (the "split") begins to appear with those in their sixties at the time of Labov's 1966 study (58). More recent work showing younger speakers again moving towards a continuous system could actually be a cyclical change in the vowel system, going back to the state found before Labov's proposed change in progress (Coggshall 2017:143). "But if the short-*a* system in NYCE did arise from a system with a single allophone or phoneme, it only adds to the mystery of how such a complex system arose in, apparently, less than a century" (Coggshall 2017:144).

4.5 Renewed Interest

After a decades long lull, a new wave of study of New York City English began in the early 2000s. The most prominent early works were the *Atlas of North American English* (Labov Ash, and Boberg) in 2006 and then Labov's 2007 paper *Transmission and Diffusion*, both of which gave extensive descriptions of the short-*a* vowel in American English, and specifically in New York City English. Both of these works continued to build on the data originally collected by Labov for SSENYC, but now with the benefit of more precise instrumental measurement of vowels. Soon after these two works came studies carried out by a new generation of scholars. They collected new corpora and used new methods to measure and categorize the variants of the vowel. This new wave of work has also expanded the scope of study to include other ethnic groups living in NYC (Becker and Coggshall 2010, Wong 2010, *inter alia*), investigating whether they participate in this part of the local vowel system.

An overarching theme of this new era of work has been whether NYCE speakers in general are moving away from the Tragerian split. The majority of studies described below have concluded that the Tragerian system is indeed decreasing in use and being replaced by the nasal split, especially in the speech of younger speakers. Work by Michael Newman, however, reveals that this effect might be a result of relying heavily on data collected only in the borough of Manhattan and not in other boroughs of NYC. Newman's work suggests that speakers in outer boroughs, such as Brooklyn and Queens, might be more conservative in retaining the Tragerian split than the Manhattanites. My work is an extension of this, metaphorically and geographically

– as my research site in Nassau County is the next step after Brooklyn and Queens when one explores outside of Manhattan.

In this section I also highlight, following Coggshall 2017, the places in which the conclusion of a waning Tragerian split may be, in part, due to misinterpretation of inherent variation in the system. I do see evidence that there are actual changes occurring in the short-*a* system for at least some speakers in some parts of the NYC region, but "this miscalculation regarding the nature of the Tragerian system allows for scholars to overstate the amount of change, especially among the oldest speakers" (Coggshall 2017:188).

4.5.1 The Atlas of North American English 2006

The Atlas of North American English (ANAE), published in 2006, documents the vowels that are relevant to regional variation in American English. ANAE describes the short-*a* vowel in depth, including its participation in several different vowel systems and shifts across the country. In ANAE, the nasal split is described as the most common system in American English dialects, and several complex systems such as New York and Philadelphia are also described. ANAE frames the raising of short-*a* as parallel to that of open-*o*, both ingliding vowels in the NYCE system with a lower counterpart (/æ/ and /ɑ/, respectively). Arguably ANAE's most influential contribution to the study of short-*a* was establishing 700Hz as the threshold to consider a token as "raised-ah" (192). There is no analysis of F2 to account for the fronting of *tense* short-*a*. This 700Hz threshold has been used by most, if not all, subsequent work as the standard for categorizing a token as raised or *tense* short-*a*.

ANAE gives a basic description of the NYCE complex split (173) but does not clearly recognize the variation in certain environments (e.g. before velar nasals) that had been noted in previous work (i.e. SSENYC, LYS, Cohen 1970). The basic description given in ANAE (173) describes short-*a* in NYCE as being *tense* in closed syllables with following voiced stops, front nasals, and voiceless fricatives, and *lax* in closed syllables with following voiceless stops, voiced fricatives, and liquids (which are referred to as "the remaining environments," though velar nasals seem to be omitted from this description). The description in ANAE also lists, but doesn't discuss in-depth, additional "phonological, grammatical and lexical conditions" affecting the *tensing* of short-*a*. These include the following:

- short-*a* is *lax* in open syllables
- inflectional boundaries don't affect the syllable structure
- weak words (i.e. function words) are *lax*
- "learned words" are *lax*
- Word-initial short-*a* is *lax* (except in "common words")
- and "avenue" is listed as a unique lexical exception

There is one point on which ANAE attempts to account for some variation. While the class of words with following voiced fricatives are first listed in the *lax* class, among the list of exceptions is a brief mention that short-*a* before voiced fricatives is "frequently *tense*" (173). The example words given for this environment, "imagine" and "magic," are both open syllables, which would make them *lax* anyway, and both specifically have the voiced affricate. While they are not clear examples for these reasons, it must still be noted that an attempt at accounting for some variation is made. Yet, to confuse the matter even further, directly following this

description of the *tense* and *lax* classes and the exceptions, ANAE includes a chart illustrating the *tensing* environments in NYCE and Philadelphia English, in which all the voiced fricatives are inside the box, indicating that they are invariably *tense* environments. This chart is reproduced below as Figure 15.

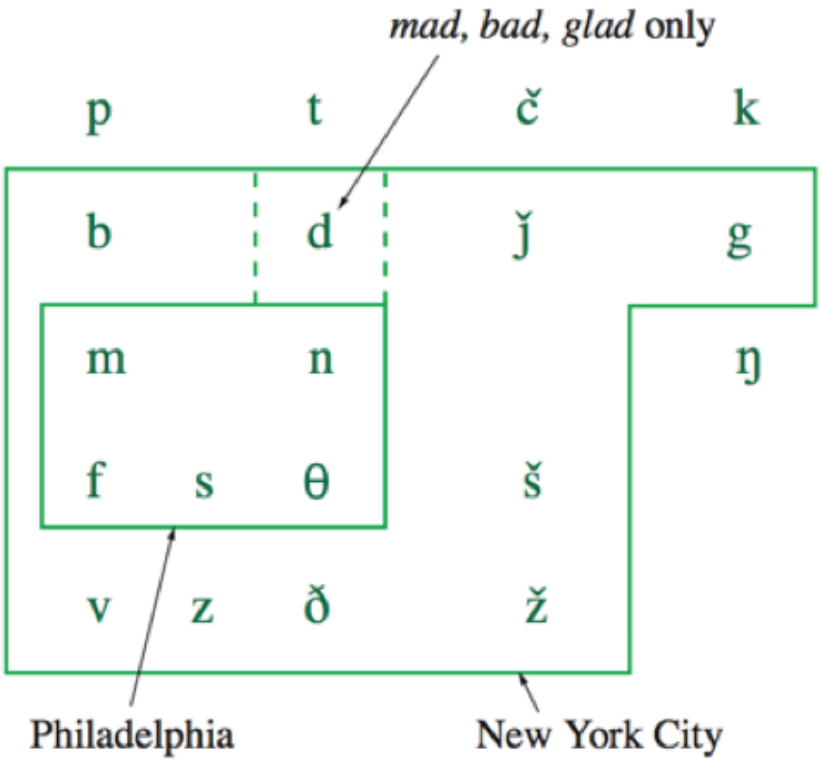


Figure 15: Following environments that condition short-*a tensing*, from the Atlas of North American English (2006:173) [5]

[5]As printed in ANAE, this chart contains an error that I have corrected in my reproduction here. I placed the /θ/ in the correct spot in the inner square of this graph, replacing the errant /æ/ that appeared there.

4.5.2 Labov 2007

Labov's 2007 article, *Transmission and Diffusion,* uses short-*a* in NYC as a case study for models of language transmission. This is the paper in which Labov's oft-cited version of the conditioning factors for the *tense* and *lax* variants appears: the following environments (shown below in Figure 16) and his seven additional constraints. While the description of the NYCE short-*a* split is not the main focus of this article, the formulation that Labov gives here has been used as the basis for almost all work on the subject that has followed (e.g. Becker and Wong 2010; Becker 2010; Coggshall and Becker 2010; Newman 2014). However, as Coggshall 2017 points out, reliance on this chart has created a problem in tracking change over time.

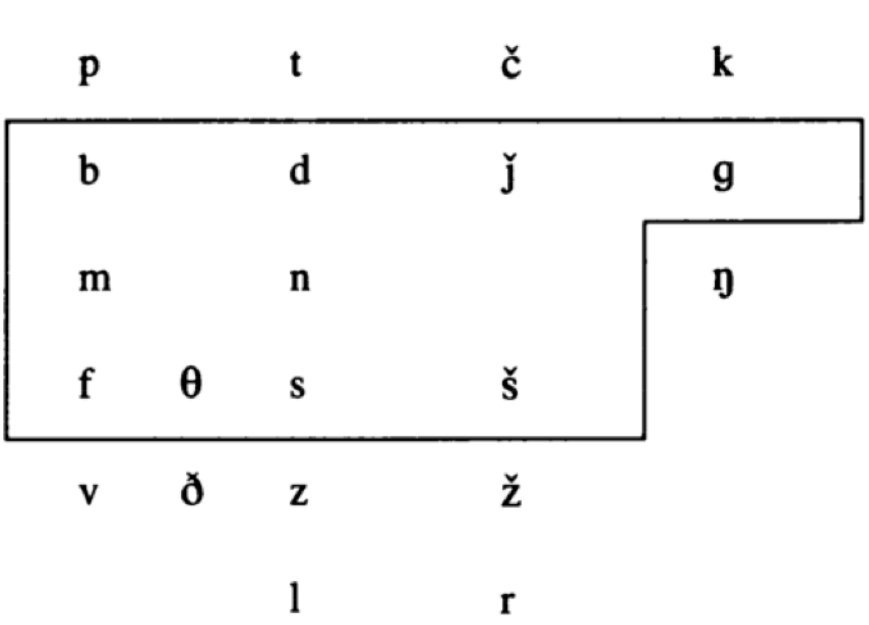


Figure 16: Following environments that condition short-*a tensing*, from (Labov 2007:54)

By formulating the phonological pattern and its various additional constraints in such absolute terms, Labov 2007 effectively erases from the discussion any consideration of the inherent variation that was noted in the foundational dialectological and early variationist studies described above. As Coggshall puts it, "Labov (2007) states the Tragerian NYCE short-*a* pattern with absolute surety, as well as surety that it remains intact today. He glosses over all of the previous data that indicate inherent variation on many levels" (154).

Labov had constructed charts similar to the one in Figure 16 before (Labov 1981, Labov et al. 2006), but one crucial detail that is introduced into the chart for the first time in Labov 2007 is that all the voiced fricatives are outside the box (meaning that they are depicted as invariably *lax* environments). In ANAE, published a year earlier, the voiced fricatives had been inside the box, among the *tense* environments (as shown above in Figure 15). Labov 2007 is the point at which the class of words with following voiced fricatives start being regarded as invariantly *lax*. The velar nasal is also on the outside of the box in Figure 16, describing it as a categorically *lax* environment instead of as a variable one, as was found in previous work.

In his 2007 article, Labov also sets forth seven additional constraints that condition short-*a* *tensing* beyond following phonological environment (354-355). Although the concepts for each have been mentioned previously in this chapter, I will describe Labov's specific formulation of each, since they are heavily referenced in the rest of the work discussed in this section. The constraints are as follows:

a) Function-word constraint: some function words such as "an, can, had" exhibit a *lax* short-*a* where a *tense* vowel appears in similarly structured content words such as "tin can" and "add." Labov 2007 applies this constraint to "function words with

simple codas" which allows him to account for the word "can't" being *tense* while its counterpart "can" is *lax*.[6]

b) Open-Syllable Constraint: vowels in an open syllable surface as *lax*. For example, the vowel in the word "plan" /plæn/ is *tense,* but the vowel is *lax* in "planet," where the word is syllabified as /ˈplæ.nət/, with the short-*a* in an open syllable.

c) Inflectional-Boundary Closing: an inflectional boundary causes the short-*a* to act as if it is in a closed syllable, rendering the vowel in, for example, "planning" *tense*, even though it is arguably in an open syllable.

d) Initial Condition: short-*a* will surface as *lax* when word-initial even if the following environment would normally condition it as *tense*, for example in the word "aspirin." Labov 2007 says that "the most common words" are exceptions to this rule, giving the two examples of "ask" and "after" being *tense*, though word-initial.

e) Abbreviations: in Labov 2007, this refers specifically to shortened forms of names, with the examples given being "Babs" and "Cass." These nicknames contain a *lax* short-*a* when a *tense* one would be expected because of the following voiced stop and voiceless fricative, respectively.

f) Lexical exceptions: this constraint accounts for a number of lexical exceptions like the word "avenue" which commonly is *tense* instead of *lax* even though it is word-

[6]This particular formulation does not account for why the word "and" is *lax* when it too sometimes has a complex coda. A perhaps better definition is that given in Labov et al. 2006, where this constraint is described as a "weak word" constraint, applying to those function words where the vowel can be reduced to schwa. This allows the constraint to apply to "and," a word in which the vowel is often reduced, but not to "can't," a word in which the vowel is not reduced (173).

initial and in an open syllable – which would both condition it as *lax*. Labov 2007 presents this as a class of "lexical exceptions" plural, that would presumably encompass other words. However, the only example given is "avenue," which had been singled out as its own lexical exception in ANAE and also other previous work.

g) Learned words: this constraint refers to *lax* vowels occurring where they would otherwise be *tense* in words that are learned later in life. Labov's examples of learned words are "alas" and "carafe."

None of these constraints are completely new; they are a collection and codification of patterns mentioned in previous work. But the list as Labov 2007 standardized it has become the default in subsequent work when describing the Tragerian split. It is what that subsequent work compares itself to in order to determine if the NYCE short-*a* system is changing (Coggshall 2017:156). As Coggshall stated, "The problem is that this clean, if baroque, version of the split does not seem to have existed for all speakers of NYCE at any time" (156).

Since Labov 2007 focuses on the nature of the split rather than the degree of æh-*tensing* that was the dependent variable in SSENYC, a consideration of stigmatization is not included. He says that "the distribution into *tense* and *lax* classes is not socially evaluated and is a general pattern in the spontaneous speech of community members, to the extent that it is not disturbed by the effects of formal observation" (Labov 2007:54, but see Sneller 2018).

4.5.3 Becker and Wong 2010

Although the ANAE and Labov 2007 precipitated the resurgence in study of short-*a* in NYC, those papers were still based on the data collected by Labov for SSENYC (1966). Becker and Wong 2010 was the first study in decades to collect new data for a large-scale empirical study of the variable, using instrumental acoustic measurement rather than impressionistic coding of tokens. They base their research questions and their assumed framework for the short-*a* split on the description given in Labov 2007, but they collected new data in order to investigate any change over time that might have taken place since Labov's data, as well as to expand that study to Chinese, Puerto Rican, and African American speakers who had previously been assumed not to take part in the local vowel system.

The study included data from 24 speakers, all of whom were native New Yorkers; twelve were White, four were African American, four were Puerto Rican, and four were Chinese. Becker and Wong found that the twelve minority speakers were not following the Tragerian split, as "all three ethnic minority groups show no significant differences in F1 means between /æ/ in the following front and velar nasal environments; further, their mean F1 of /æ/ in the Labovian *tensing* environments [was] not significantly different from that of the Labovian *lax* environments" (19).

Becker and Wong also found age grading amongst their white speakers, with the older speakers having Tragerian splits, the middle-aged speakers having a more continuous intermediary system, and the youngest speakers approaching full nasal-split systems. They cite *tensing* before velar nasals as the prime evidence of middle-aged speakers not having *as faithful* a Tragerian system as the older speakers. But as this is one of the classes that Trager himself

mentioned as having variation back in the 1930s, there might not actually be as fine grained a difference between the older and middle-aged white speakers as Becker and Wong surmised from their data. It still remains, however, that their youngest white speakers had "far more simplified systems" with few of the additional constraints that characterize the Tragerian split (18), lending support for their findings of apparent time change.

Figure 17 below (from Becker and Wong 2010), which ranks the height of each class of words (by phonological environment) for each group of speakers, shows that the difference between the F1 of the lowest *tense* class and highest *lax* class is 82 Hz for older white speakers, 14 Hz for middle-aged speakers, and 12 Hz for the youngest white speakers. So while the middle-aged and younger white speakers still have a split according to Becker and Wong, their systems are more continuous, with less of a difference in height between their *tense* and *lax* classes (18). This figure also shows that the youngest speakers of all ethnicities in their study raised short-*a* before both front and velar nasals.

	White Old	White Middle	White Young	African American	Puerto Rican	Chinese
Highest /æ/	Vl. Fricative (*507hz*)	Front Nasal (*562 Hz*)	Front Nasal (*574 Hz*)	Front Nasal (*531 Hz*)	Velar Nasal (*566 Hz*)	Velar Nasal (*543 Hz*)
	Front Nasal (*527 Hz*)	Vl. Fricative (*569 Hz*)	Velar Nasal (*623 Hz*)	Velar Nasal (*541 Hz*)	Front Nasal (*627 Hz*)	Front Nasal (*568 Hz*)
	Vd. Stop (*561 Hz*)	Vd. Stop (*598 Hz*)	Vl. Fricative (*634 Hz*)	Vd. Stop (*643 Hz*)	Liquids (*649 Hz*)	Vd. Fricative (*662 Hz*)
	Liquids (*643 Hz*)	Velar Nasal (*612 Hz*)	Vd. Stop (*648 Hz*)	Vl. Stop (651 Hz)	Vl. Stop (*656 Hz*)	Vd. Stop (*679 Hz*)
	Velar Nasal (*644 Hz*)	Vd. Fricative (*632 Hz*)	Vd. Fricative (*660 Hz*)	Liquids (*681 Hz*)	Vd. Stop (*672 Hz*)	Liquids (*700 Hz*)
Lowest /æ/	Vl. Stop (*677 Hz*)	Vl. Stop (*668 Hz*)	Vl. Stop (*695 Hz*)	Vl. Fricative (*669 Hz*)	Vd. Fricative (*689 Hz*)	Vl. Stop (*706 Hz*)
	Vd. Fricative (*686 Hz*)	Liquids (*688 Hz*)	Liquids (*731 Hz*)	Vd. Fricative (*700 Hz*)	Vl. Fricative (*691 Hz*)	Vl. Fricative (*714 Hz*)

Figure 17: Gray boxes represent word classes described in Labov 2007 as being *lax*, while white boxes represented *tensing* environments, from Becker and Wong (2010:10)

To further investigate which constraints were being used by each of their speaker groups, Becker and Wong did a linear regression analysis with F1 as the dependent variable, and the Labovian constraints of Following Environment, Syllable Structure, Function Words, Lexical Exceptions, and Word Initial tokens (which are listed in that order in the first column of the model) as the independent variables. The results of this regression analysis for F1 is presented below in Figure 18.

	White Old	White Middle	White Young	African American	Chinese	Puerto Rican
Following Environment	$p < .0001$	$p < .0001$	$p < .0001$	$p < .0001$	$p < .0001$	$p < .0001$
Syllable	$p < .0001$	$p < .0001$	$p < .0001$	$p = .0020$	$p = .0200$	
Function	$p < .0001$	$p = .0030$				
Exception	$p < .0001$	$p = .0330$				
Word Initial	$p = .0020$					

Figure 18: Significant factors on F1 for the different ethnic groups, from Becker and Wong (2010:17)

The analysis found evidence of an Open Syllable Constraint being a significant conditioning factor for all groups other than Puerto Ricans; this result includes white speakers of all ages, not just the older ones determined to have a Tragerian split (17). They also found that following environment was significant for all ethnic groups, but the way in which different classes of following environment affect the patterning of short-*a* allophones was not the same for each group.

Becker and Wong concluded that the short-*a* system in NYCE is "losing its complex conditioning over time" for white speakers, and that young native New Yorkers of ethnic minority backgrounds (Chinese, Puerto Rican, and African American) who speak English natively do not produce the traditional NYCE split, but instead produce a nasal tensing system" (12). While their observation of a change in progress may be sound, the transitional stages that they identify are up for reinterpretation given Coggshall's important reframing of the literature.

4.5.4 Becker 2010

Becker 2010 revisited the Lower East Side of Manhattan forty years after Labov's SSENYC. Using data from interviews with 64 speakers native to the neighborhood, Becker does a thorough analysis of short-*a* and THOUGHT-raising in NYCE. She found that the change in progress that Labov had predicted, a continued pattern of /æ/-raising, was not borne out. Instead she found that the NYC complex split seemed to be diminishing among the younger speakers in her sample, who were more likely to have continuous or nasal-split systems. Expanding on Labov's 1966 work, Becker includes in her sample African American, Asian, and Latino participants, who she finds do participate to some extent in the NYCE short-*a* split.

Becker's speakers are divided into three age groups: Older (born between 1924-1951, N=7), Middle-Aged (born between 1952-1972, N=21), and Young (born between 1974-1990, N=6) (169). Her sample includes speakers of five ethnicities: White (which she breaks into Jewish and non-Jewish), African American, Asian American, and Latino.

Becker categorized her speakers into five different patterns of short-*a* that she calls Labovian, Labovian: Transitional, Continuous, Nasal, and Nasal: Unclear. (What Becker calls "Labovian" is what I call "Tragerian" throughout this dissertation.) She finds "true" Tragerian splits only in speakers born before 1971 (13 of 21), and that group includes one Asian American and one African American exhibiting the Tragerian system (172), which was a new finding. Labovian: Transitional is described as close to a Tragerian split, but with some variation (that is not exactly enumerated). Becker puts six speakers in this category, who are either Older or Middle Aged and are a diverse ethnic cohort (1 Jewish speaker, 2 White speakers, 2 Latino speakers, and 1 African American speaker) (181). Becker sees this transitional group as

indicative of a change in progress away from the Tragerian system (172). It is possible, though, that Becker's distinction between Labovian and Labovian: Transitional is based heavily on environments (e.g. velar nasal) that might have always been variable. This perspective would suggest that her "transitional" speakers might, in fact, be better categorized as having full Labovian systems.

The Continuous category accounts for nineteen of Becker's speakers, who are mainly younger and not white. The nasal-split group is the most ethnically diverse group, including five white speakers (including Jewish and non-Jewish), 2 African American speakers, 2 Asian American Speakers, and 2 Latino speakers. Becker also does not precisely describe what factors differentiate her Nasal category and her Nasal: Unclear category, but it seems to be a mixture between a nasal split and a continuous system. Crucially it is not Tragerian. The Nasal: Unclear group is mainly Middle-Aged and includes fifteen speakers (181). 20% of Becker's speakers have a Tragerian system (or close to it), while nearly half have a continuous or nasal system (182).

Becker tests whether or not her speakers, including those with nasal splits, maintain "remnants" of the Tragerian system, that is, adhere to any of the additional constraints beyond phonological environment (192). By doing t-tests on the average F1 and F2 values for each speaker in open versus closed syllables, she finds that the Open Syllable Constraint holds for all of her Older and Middle-Aged speakers (186). She finds, however, that speakers in her corpus do not seem to have the Function Word Constraint, with the auxiliary use of the word "can" and the conjunction "and" being more fronted by her speakers than the *lax*-environment function words, showing a nasal split even among function words. With regard to the Initial Condition, Becker

has mixed results (193). She finds that there is still a tendency to produce any word-initial short-*a* as *lax*, but that this is increasingly not true when the following segment is a front nasal. She also found that the two exceptions mentioned by Labov "ask" and "after" are not *tensed* by her speakers (194).

Finally, Becker looks at the famous lexical exception "avenue." She finds a lot of variation among her speakers, even the oldest ones with the most intact Tragerian systems (196). She does find "avenue" to be significantly different (i.e. more *tense*) than other words where the vowel precedes a voiced fricative, and so concludes that this lexical exception is maintained in her corpus.[7] Over-all, Becker concludes that "the speakers in this corpus do not abide by [the Labovian additional conditions] and confirm the picture of a change over time in which the Labovian system is not produced by younger speakers of any ethnicity" (192).

Becker also performed a regression analysis with F1 of short-*a* production as the dependent variable and age and ethnicity as independent variables. She found that adherence to the Tragerian split varies significantly by both of these social variables. Front nasals most favor *tensing* for all of the ethnic groups, and velar nasals are the highest *lax* class for white speakers (both Jewish and non-Jewish). Becker sees this positioning of the velar nasal class as evidence of a move towards a nasal system (202).

Becker uses a series of ANOVA tests (based on those in Becker and Wong 2010), in which she reduces the complexity of the groups of following environments down to just four:

[7]Becker does not test Inflectional Boundary Closing or the Abbreviation constraint because she does not have enough tokens to do so. Additionally, she doesn't look at learned words because she finds them "too subjective" (165-166).

(front nasal, velar nasal, all other Labovian *tense* and all other Labovian *lax*) in order to more clearly illustrate patterns between ethnic groups and age cohorts. She finds that white speakers display a drastic change over apparent time for both F1 (Figure 19 below) and F2 (Figure 20 below), with Labovian *tense* class of words getting progressively more *lax* and the velar nasal class getting more *tense* as one moves along the x-axis from old to middle to young speakers. Front nasals and Labovian *lax* stay more or less stable as the most and least *tense* respectively (the Labovian *tense* does drop below the Labovian *lax* average a little bit for the young speakers, but within the margin of error indicated by the ellipses).

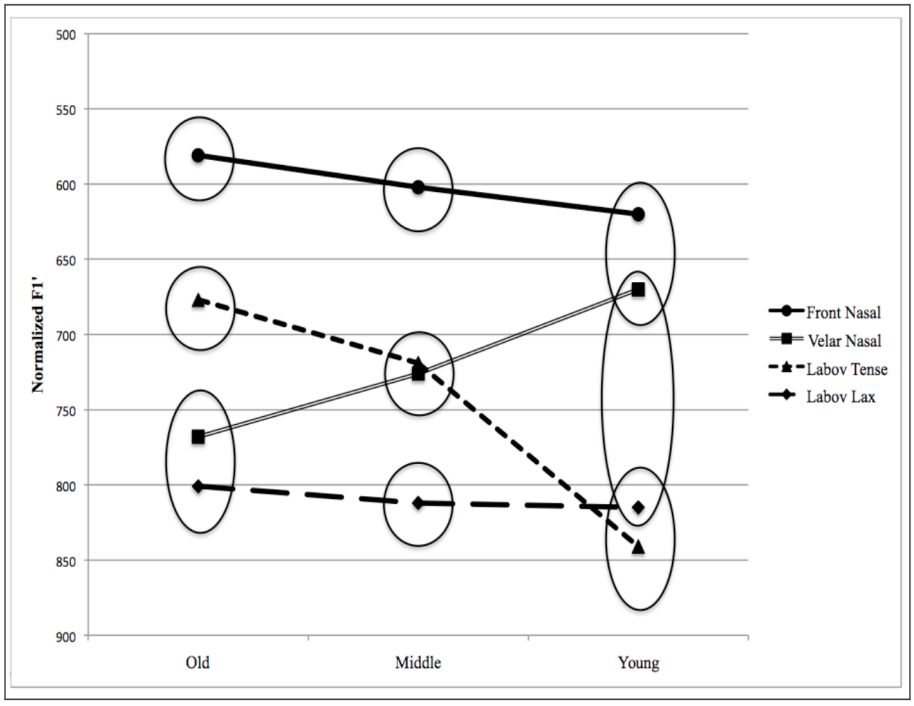


Figure 19: Mean Normalized F1 of /ae/ for white speakers by age, from Becker (2010)

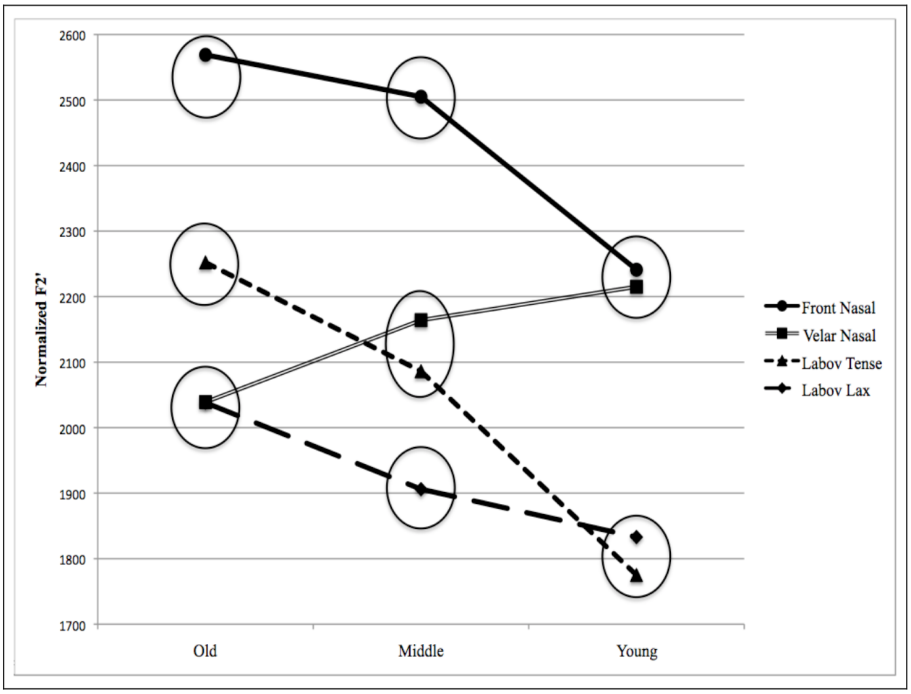


Figure 20: Mean Normalized F2 of /ae/ for white speakers by age, from Becker (2010)

In sum, Becker (2010) finds a shift in the whole community from a Tragerian short-*a* split towards a nasal split. Her youngest speakers of all ethnicities behaved more-or-less the same, although she found Tragerian systems among older speakers from ethnic communities that had previously been thought not to participate in the local vowel system (Labov 1966, Cohen 1970). Becker suggests that measuring "the gap between the vowel height for words with nasal codas and words with other codas" (188) might be a better measure of the change towards a nasal system.

4.5.5 Newman 2011, 2014

Newman's work is some of the only research looking primarily at the outer boroughs (e.g. Brooklyn and Queens) of New York City, whereas most of the previous work on short-*a* has focused solely on Manhattan. This makes it a bridge between my work and the Manhattan studies, since it is possible that the changes being observed may correlate with distance from the center of the city. Crucially, Newman's 2011 and 2014 work present evidence of the retention of the Tragerian split by younger speakers in New York City that were found to be almost non-existent (Becker 2010) in younger Manhattan speakers.

The sample of speakers in Newman 2011 is four white families from the Bronx and Brooklyn, with birth years ranging from 1925-2002. He talks about them in terms of their generational status in the family, with age cohorts corresponding to old, middle, and young. Newman categorizes each of his speakers as having an *intact* [Tragerian] system, an *intact* [Tragerian] + /ŋ/ system (characterized by raising in environments preceding velar nasals in addition to the other traditional *tensing* environments), or a *mixed system* (described by Newman as "all tautosyllabic nasals *tensed* plus some traditional NYCE tensing environments"). No speakers with nasal splits were found in this study. He does find, though, that none of his young speakers have a completely *intact* system. Newman characterizes four out of eleven of the young speakers as having *mostly intact* systems. Since this categorization is based on the velar nasal word class, which has been shown to always have had some variation, this means that the majority of the speakers in this study, including the young speakers, do in fact have Tragerian systems. Newman does still conclude that a shift away from the Tragerian system over time is

occurring, and he creates an implicational scale listing features of the classic Tragerian split in the order that they are most likely to be lost by younger speakers.

In his implicational scale, *tensing* of velar nasals is the feature listed as most likely to change in the process of losing the Tragerian short-*a* split, which he finds to be a universal feature among all of his speakers born after 1972. That is followed in order by the *tensing* of open syllables, a *lax* short-*a* before voiceless fricatives, a *lax* short-*a* before voiced stops, and finally, *laxing* of the word "after."

Newman 2011 presents evidence of young white speakers in New York City, specifically the outer boroughs, exhibiting a higher rate of retention of Tragerian features than Becker found in the Lower East Side. He highlights the velar nasal as the most prominent indicator of and the first step in the process of change, as Becker did. Slightly diverging from Becker, though, he identifies the Word Initial Constraint exception "after" as being the most resistant to *tensing*, whereas Becker found the lexical exception "avenue" to be the most resistant.

The subjects in Newman 2014 are the *BQ16,* who are sixteen "mostly young" New Yorkers of varying ethnicities from the boroughs of Brooklyn and Queens (Newman 2014:5). In this study, Newman finds a more dramatic retention of the classic Tragerian system by young speakers, most of whom were born in the 1980s (58-61).

In his book, Newman offers a revision of the much-cited figure from Labov 2007 that adds more acknowledgement of variable environments. This figure (Figure 21 below) emulates the similar figure from Labov 2007 with some changes. One superficial change is the use of IPA symbols to represent each sound rather than the APA symbols that Labov uses. More substantially, though, is the addition of dotted lines forming a dotted box around the voiced

fricatives, indicating that short-*a* before voiced fricatives is variably *tense* or *lax* (as mentioned by Cohen 1970 about "jazz" and "razz"). This is a significant advancement in putting recognition of variation back into the model. Despite having previously deemed velar nasals to be on the top of his implicational scale, he still leaves it decidedly outside of the box in the "categorically *lax*" section of his chart.

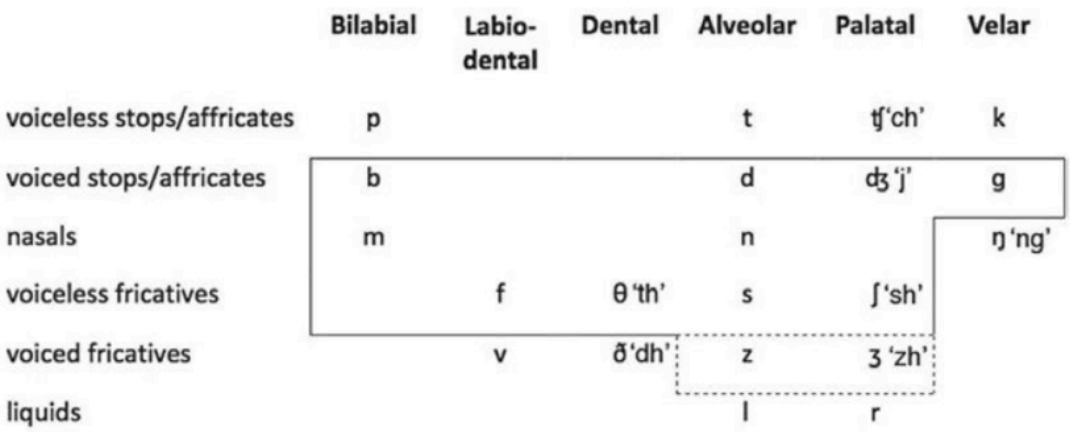


Figure 21: Following environments that condition short-*a tensing*, from Newman (2014:55)

Newman concludes that the data from both of these studies do not refute the trend found by Becker and others that the short-*a* system is simplifying and becoming less common among young white New Yorkers, but that it "suggest[s] that this decline may be happening much more slowly outside the areas they surveyed" (2014: 61).

4.5.6 Newlin-Łukowicz 2015

Newlin-Łukowicz looked at short-*a* as one of the variables in her study of the ethnic and place identities of Polish immigrants in New York City. For her analysis of short-*a*, she relies on Labov 2007 for her division of tokens into *tense* and *lax* classes for what she terms a "classic" split (245). She does not consider the variation before voiced fricatives and affricates that Labov does mention in 2007.

Newlin-Łukowicz used data from reading passages recorded by 35 Polish Americans, her target subjects, and 10 native New Yorkers that she used as a control group. Her Polish-Americans are broken up into age groups of "first generation" and "second generation," while the New Yorkers are divided into "old" and "young." She found that her older non-Polish speakers have a transitional or continuous (Becker 2010) system, and that all the speakers in her three other groups have nasal systems (253). From this, she concludes that "NYCE is losing a number of rules or constraints on raising, but it is also generalizing raising before front nasals to velar nasals" (2016a: 293).

Newlin-Łukowicz's speaker with the most Tragerian system was Cassie, an older speaker from the control group. Among the tokens that are marked "out of place" for Cassie, most can be accounted for by variability recorded in earlier work that is dropped from Labov 2007 (which, as noted, Newlin-Łukowicz relies on her for her categorizations).

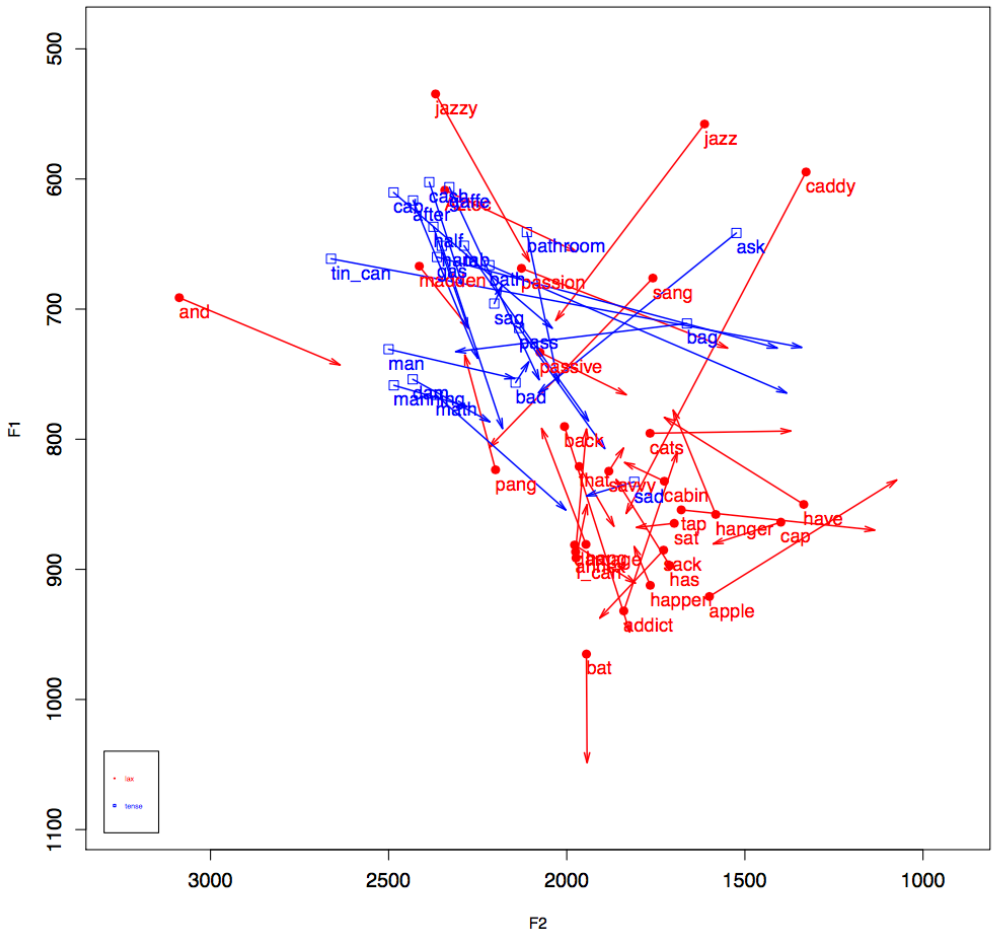


Figure 22: Cassie's ("older" control) short-*a* tokens, assuming the classic split, from Newlin-Łukowicz (2015:256)

Cassie's "out of place" tokens are "and, jazzy, jazz, Aztec, passive, passion, madden, caddy, sang, and pang," all of which Cassie pronounces as *tense* instead of *lax*. "Jazz," "jazzy," and "Aztec" all have following voiced fricatives, which is a previously noted variable environment (although "Aztec" could suggest a possible loss of the word-initial constraint). "Sang" and "pang" precede velar nasals, the most commonly noted variable environment in past work.

Coggshall (2017), discussed below, identifies “passive” and “passion” as examples of the variable nature of intervocalic /s/ and /ʃ/, but I propose in the next chapter that these words could also be accounted for by boundary closing by suffixes other than specifically inflectional affixes.

Newlin-Łukowicz found that among her native New Yorker control speakers, the older speakers had a more intact Tragerian short-*a* system than the younger speakers. The one exception to this pattern in her data is particularly relevant to the current analysis, as the speaker is from Long Island. Newlin-Łukowicz’s subject Jessica, whose short-*a* vowel plot is reproduced below in Figure 23, is the only young native New Yorker who seems to have a Tragerian split, and she is also the only one reported as being solely from Long Island and not from a borough of New York City proper.

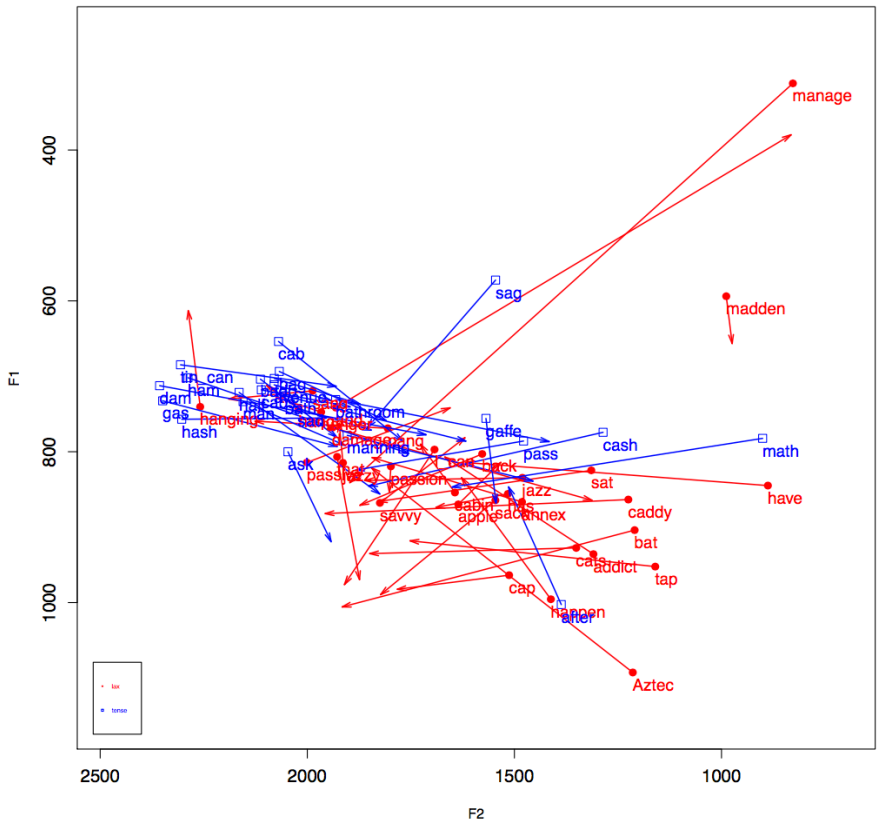


Figure 23: Jessica's ("younger" control) short-*a* tokens, assuming the classic split, from Newlin-Łukowicz (2015:262)

4.6 Coggshall 2017: Reframing the Split

Throughout this chapter I have quoted Coggshall 2017 more than any other work. This is because her recent dissertation on short-*a* in Jersey City includes a comprehensive review of the entire body of previous work on short-*a* in New York City English. This review has been an invaluable resource. What I see as Coggshall's primary contribution to the ongoing study of short-*a* is that she shed new light on underlying variation in the system, demonstrating that it had

been noted by researchers from the earliest documentation of the split, and that it is only relatively recently that researchers working off Labov 2007 have recast these patterns as absolute. Coggshall warns that it is imperative to take the Tragerian system as laid out in Labov 2007 with "a heavy amount of skepticism." She posits that the "majority of NYCE speakers differ from this strict set of rules in *predictable* ways" (emphasis in the original), and further that "the variable nature is part of an accurate description of the NYCE short-*a* system and not an aberration or a change in progress" (257). Coggshall's analysis calls into question some of the specific proposed indications of a declining short-*a* system, particularly *tensing* before the velar nasal, but she does not argue that there is no change happening at all.

From her extensive review of the literature, Coggshall identifies nine areas where inherent variability has been observed before and so should be examined more closely by those engaged in research on the short-*a* split in New York City English. The environments on Coggshall's list are the following:

1. before voiced fricatives
2. before voiced affricates
3. before intervocalic [ʃ] (potentially also before intervocalic [s])
4. before [ŋ]
5. in open syllables with following nasals, voiced stops, and voiceless fricatives
6. before [nj]
7. in weak words, particularly with following nasals
8. before a consonant at an inflectional boundary
9. word-initially (258)

Coggshall constructs a modified version of the chart originally published in Labov 2007, and then updated in Newman 2014, which can be seen in Figure 24. Within the solid lines are the environments that condition a *tense* short-*a,* the inherently variable environments are marked by dotted lines, and all others are *lax* environments.

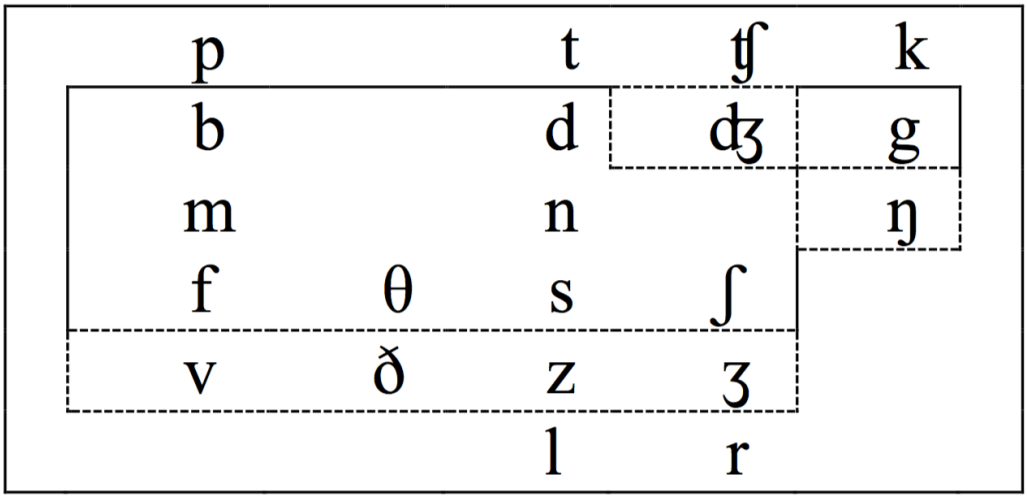


Figure 24: Following environment rules for the Tragerian short-*a* split, from Coggshall (2017:258)

In Coggshall's version of this iconic representation of the short-*a* split, there are far more dotted lines indicating variability than in either Labov or Newman's versions. Inside the solid box, as *tense* environments, are voiced stops, front nasals, and voiceless fricatives, while totally outside the box, as *lax* environments, are voiceless stops and affricates, as well as /l/ and /r/. In the variable dotted-line zone lie the voiced affricate, the velar nasal, and voiced fricatives.[8]

[8]It is my interpretation based on the context of Coggshall's work that /g/ in Figure 24 is in a solid line box, although it is visually unclear due to its neighboring phonemes (/ʤ/ and /ŋ/) having dotted lines.

In addition to reframing the previous body of literature on short-*a* in the way described above, Coggshall 2017 also adds new data from Jersey City as evidence of the state of the split in an area adjacent to but outside of New York City. Coggshall relies heavily on the methodology of Cohen (described in section 4.2.2 of this chapter) for her work, using his word classes as a basis for her own set. Previous research differs on how finely categorized the tokens in the analysis are. Cohen has the most fine-grained categorization with 28 categories, and SSENYC is on the other end of this spectrum with the coarsest categorization of just three categories: *tense*, *variable*, and *lax*. To include all the areas of variation that she has identified, Coggshall builds her analysis on Cohen's but adds even more categories to account for additional sites of possible variation.

Coggshall's analysis has 33 classes of short-*a* words. These classes account for all the possible combinations of the following phonological environments with open or closed syllables, word-initialness, inflectional morpheme boundaries, and function words. She introduces these categories in Tables 27-36 on pages 260-263. I give a condensed overview of these classes in Tables 4 and 5 below, organized in such a way as to illustrate which conditions each class represents. The below charts account for 29 of Coggshall's 33 classes. The final four are *Shortenings* or nicknames (BABS), *Learned Words* (LAD), *Lexical Exceptions* (AVENUE), and the function word exception "can't."

Following Environment	Closed Syllable	Open Syllable	Inflectional Morpheme Boundary
Voiceless Stops and Affricates	BAT	HAPPY	BATTING
Voiced Stops	BAD	TABBY	PADDING
Voiceless Fricatives	BASH	PASSION (only /ʃ/) PASSAGE (only /s/) TAFFY (all others)	BASHING
Voiced Fricatives	JAZZ	DAZZLE	HALVING
Voiced Affricate	BADGE	BADGER	
Front Nasal	BAN	PLANET CANYON	BANNING
Velar Nasal	BANG	HANGAR	BANGING

Table 4: Coggshall's Word Classes involving Closed and Open Syllables and Inflectional Morpheme Boundaries.

Note that there are three separate classes of words that are open syllables with following voiceless fricatives. The default class of this combination is TAFFY, while PASSION and PASSAGE are added to account for possible variation in intervocalic /s/ and / ʃ/. In addition, Coggshall has a separate class for words like CANYON, which I have placed in the column for open syllables for reasons that I will describe in more detail in Chapter 5. The box that would contain a following voice affricate with an inflectional morpheme boundary is grayed out, as Coggshall does not have a class to represent this condition; the number of tokens that fit this description are vanishingly rare.

Following Environment	Function Words	Word Initial
Following *Tense*	HAD (non-nasal)	ADD
Following *Lax*	THAT	APP
Following Nasal	AM	
Following "Variable"		ASTHMA

Table 5: Coggshall's Word Classes for Function Words and Word Initial Conditions

In terms of classes for function words and the word initial condition, Coggshall condenses her possible following environments down to three: following nasal, following Tragerian *tense* (other than nasal) and following Tragerian *lax*. Coggshall does not include a word class that would differentiate between word-initial tokens of short-*a* that are followed by a nasal or another Tragerian *tense* environment, so that box is grayed out in Table 5.

By separating out all 33 of these conditions instead of lumping them together, Coggshall is able to shed light on where exactly variation can be found, and then connect those environments to where it either has or has not been noted in the past. This helps distinguish between two kinds of variation found in contemporary studies, that which has been present all along as opposed to that which potentially represents changes in progress within the short-*a* system.

4.6.1 Coggshall Results

Coggshall's data is from sociolinguistic interviews with twenty residents of Jersey City, born between 1918 and 1984. Of her speakers, twelve are white, four are black, three are Filipino, and one is Latino. Many of them are the children of immigrants or immigrants

themselves who arrived in the United States at a very young age (72-73). Coggshall notes that since she is using such fine-grained categories, she does not have data for all of the categories from any one speaker, but she was able to cover most of them through reading passages and word list tasks (259). She includes all stressed words with one or two syllables, and then selects two tokens of each lexical item. This results in 1898 tokens from interviews and 1031 from reading tasks (these include a reading passage, a word list, and a set of minimal pairs). She codes each word into one of her 33 classes introduced above in Tables 4 and 5 (285).

What emerges from this meticulous analysis is an "over-arching pattern that white speakers have Tragerian short-*a* systems and black and Filipino speakers have nasal systems." (291). She categorizes each speaker in her study into one of three categories according to the short-*a* system that they exhibit: Tragerian, Nasal, or No Split. She did this by categorizing each word class for each speaker as being *tense*, *lax*, variable, or "borderline." Below in Figure 25 is a table in which Coggshall lists each participant and their categorization for each closed syllable word class. She also includes analogous charts for the open syllable classes, inflectional morpheme boundary classes, function word classes, word-initial classes, and a final table with the remaining exceptional classes (Tables 43-48, pages 292-303).

Speaker	BAT	BAD	BAN	BANG	BASH	JAZZ	BADGE
Camille	lax	lax	tense	tense	lax	lax	--
Claudette	lax	lax	tense	tense	lax	[lax]	--
Terrence	lax	[lax]	tense	lax	lax	[tense]	--
Lynn	lax	lax	tense	[tense]	lax	[lax]	
Daniel	lax	lax	tense	variable	lax	lax	border
Jeff	lax	lax	tense	tense	lax	lax	--
Fely	lax	lax	tense	tense	lax	lax	--
Will	lax	border	tense	tense	variable	lax	--
Philip	lax	tense	tense	lax	variable	variable	--
Martha	lax	tense	tense	lax	variable	--	--
Joan	lax	tense	tense	[border]	variable	--	--
Patrick	lax	tense	tense	variable	tense	lax	--
Ed	lax	tense	tense	lax	tense	variable	--
Anthony	lax	border	tense	lax	border	variable	--
Don	lax	tense	tense	variable	tense	variable	--
Tommy	variable	tense	tense	variable	tense	variable	[tense]
Charlie	lax	tense	tense	lax	tense	[border]	--
Dorothy	lax	tense	tense	lax	tense	--	--
Anna	lax	tense	tense	[lax]	tense	--	--

Figure 25: Short-*a* for all speakers in closed syllables, from Coggshall (2017:292)

Coggshall categorizes most of her speakers into just two categories, Tragerian or Nasal. She does not include other intermediate classes as other researches have done (e.g. Becker 2010 and Newman 2014), because she recognizes a higher tolerance of variation within each category. In fact, she says "the only real overarching theme to the results for these speakers is one of variation" (312). Among her Tragerian speakers, Coggshall finds variation in the BANG word class (following velar nasal in a closed syllable) as many others have, but she finds no correlation by age that would indicate it is part of a change in progress (310). This supports her hypothesis that this environment has always included some variability. She also finds that some of her speakers have variable pronunciations in the JAZZ word class (following voiced fricative in

a closed syllable), another environment in which she was expecting to find variation. Some of the speakers with the Tragerian system are also categorized as *variable* for the PLANET word class (following front nasal in an open syllable), but none of them are invariably *tense* in this environment. Finally, four speakers with the Tragerian system (all young and white) have variable pronunciation of short-*a* in the BASH word class (following voiceless fricative in a closed syllable). She suggests that this might be the next "basic tenet" of the Tragerian system to be weakening (308).

Coggshall also notes variation within the group of speakers that she categorizes as having a nasal split. There is variability in the BANG word class here as well, with some nasal speakers having variable pronunciations of short-*a* before the velar nasal. As Becker did, Coggshall also finds that some of her nasal speakers exhibit the Open Syllable Constraint, producing *lax* short-*a* sometimes before front nasals when it occurs in an open syllable (294). The two speakers with the nasal system who have the Open Syllable Constraint also show the effect of morpheme boundary closing in the BANNING word class. Coggshall also finds that some nasal speakers *tense* the short-*a* in "avenue," but she suggests that this might be a special exception that has taken on an indexical function for residents of the New York City area (306).

Coggshall additionally presents an implicational scale that ranks word classes in how likely they are to be variable in a Tragerian speaker's system. Newman 2011's scale aimed to represent change over time, ranking features of the split as more or less likely to be lost. Contrastingly, Coggshall's is focused on word classes that appear to have been variable for the history of the Tragerian system, which happen to correspond to the classes of words that researchers have often offered as examples of a changing or diminishing short-*a* split. In Figure

26 below, Coggshall's implicational scale lists each of the speakers with a Tragerian system and their pronunciations of the lexical exception "avenue" as well as the PLANET, PADDING, BASHING, AM, and ADD word classes. In this scale, a + indicates that the speaker consistently has the classic Tragerian variant for that word class, and a – indicates that they consistently do not. A +/- indicates that the speaker varies between *tense* and *lax*, and the ~ indicates that the speaker's short-*a* is borderline in that environment. Parentheses in the table indicate where the data errs from the scale, and if a speaker did not have any tokens of a given category, this is marked as n/a.

		AVENUE	BANG	PLANET	PADDING BASHING	AM	ADD
	Speaker:						
1955	Martha	+	+	+	n/a	+	-
1918	Anna	+	+	+	+	+/-	-
1941	Ed	+	+	+	+	~	-
1931	Charlie	+	+	+	+	-	-
1937	Don	+	(+/-)	+	+	-	-
1956	Philip	+	+	+	+/-	(+/-)	-
1937	Anthony	+	+	+/-	+/-	(+)	-
1921	Dorothy	+	+	+/-	+/-	(~)	-
1934	Tommy	+	+/-	+/-	n/a	-	-
1951	Joan	+	~	n/a	n/a	-	-
1963	Will	+	-	+/-	+/-	-	-
1950	Patrick	~	(+/-)	(+)	-	(~)	-

Figure 26: Implicational scaling for speakers with a Tragerian system, from Coggshall (2017:318)

Coggshall's implicational scale shows that the lexical exception "avenue" and the BANG word class are actually the least variable among the speakers in her study with the Tragerian system, despite being reported by Becker and Newman as the most volatile parts of the system. It is the function word constraint and the word initial condition (AM and ADD) where these speakers are

most likely to not follow the Tragerian split. The Open Syllable Constraint (PLANET) and Morpheme Boundary Closing (PADDING and BASHING) are the most likely to be variable among these speakers.

Going a step further, Coggshall proposes an explanation for why certain classes are more likely to be inherently variable and others are more likely to be stable. While many have attempted to *describe* the short-*a* split (as illustrated by this "short" history), this is the first attempt to *explain* it that I am aware of. The split makes little phonological and phonetic sense. Some propose that this is because it is phonemic (e.g. Trager 1940, Hubbell 1950), but Coggshall believes that it is a pattern of frequency and that Exemplar Theory may explain it (265). In Exemplar Theory (e.g. Bybee 2001, Pierrehumbert 2001), every token of a given linguistic form that one hears is stored in a neural cloud, and the more times that you hear something, the bigger the cloud grows. A bigger cloud allows for more detailed information about that form to be stored and extracted. Coggshall hypothesizes that environments that frequently follow short-*a*, and so have larger exemplar clouds, are more reliably *tense* or *lax* for speakers, but those that are infrequent are the ones that tend to be variable (266).

Coggshall tested her frequency hypothesis by finding the frequencies for words that have short-*a* followed by each phonological category in several corpora: MRC Psycholinguistics Database, CELEX, and then in her own data. In all three corpora, the frequency data largely fits her hypothesis. The uniformly *lax* classes (BAT) and uniformly *tense* classes (BAN, BASH, and BAD) are more common than the inherently variable classes BADGE and JAZZ. However, she found that the velar nasal class is not actually that infrequent, not correlating with how variable it has been described to be. BANG comes in at slightly more frequent than the BAD word class in all

three corpora. It has fewer tokens than [m, n], making it reasonable to consider it a separate category from front nasals, but it does not have as few as the other variable classes (269-270). The variability of the velar nasal class cannot be explained by frequency alone. However, one contributing factor is that among velar nasal tokens one lexical item accounted for the preponderance of data: the word "thanks." Because "thanks" is very common in everyday speech, one would expect it to have an exceptionally large exemplar cloud when compared to the rest of the words in the BANG class of words. Coggshall proposes that the variability of pre-velar nasal short-*a* might also be impacted by the ambiguous and exceptional nature of [ŋ] in English (274).

Coggshall also believes that there is a frequency explanation for the different treatments of the two affricates. Labov (2007:354) analyzes the two affricates differently: the voiced affricate is treated like a voiced stop, and the voiceless affricate is categorized with the voiceless fricatives. Coggshall proposes that the difference in behavior between these two phonemes with the same manner and place of articulation could also be due to frequency of short-*a* words than contain them. BADGE is the class of words with voiced fricatives, but it is difficult to think of many other closed syllable words that fit the bill besides "badge." Coggshall consulted several dictionaries and native speakers and could only come up with a handful of BADGE words, and half of them were shortenings of longer words (274). I cannot do any better in finding words to fit this category and agree that it is exceedingly infrequent. In contrast, the voiceless affricate is in the top half of following environments in regard to frequency in all three corpora.

Coggshall's dissertation dedicates as much space to reviewing the literature on short-*a* as it does to presenting new findings. This is for a very good reason, as it is essential to understand

the past analyses of this complex system in order to study it as it is now. Her in-depth review also reveals how problematic it is to rely on a single published version of the Tragerian short-*a* system to make a case for current changes in progress (338). In Labov 2007, the only site of variation that is acknowledged is that of before voiced fricatives, but forty years earlier in Labov 1972, variation had been noted in several other environments, including before velar nasals. The data presented in recent studies (e.g. Becker and Wong 2009; Becker 2010; Newman 2011, 2014; Olivo 2013, Newlin-Łukowicz 2015) show that much of this variation appears to have been passed down to speakers today. But by mainly comparing their results only to Labov 2007, they sometimes erroneously use this variation to assert that the Tragerian system is being lost rather than as evidence of the maintenance of a complex system with inherent variation.

4.7 Conclusion

This chapter has given an overview on the long history of study of short-*a* in New York City, and a survey of the many and varied methods of categorization and analysis that have been used. The main narrative that I want to highlight is one that comes from Coggshall 2017. Throughout much of the early study of this vowel system, variation had been noted in certain environments, notably before velar nasals and voiced fricatives. These are low frequency environments and seem to not have ever been invariably *tense* or *lax* for speakers. However, many recent studies use Labov 2007 as a starting point for short-*a* analysis, which is problematic because in that paper, Labov omitted almost all of the consideration of inherent variation in the system, stating the conditioning properties of New York's short-*a* split in absolute terms. In my

analysis, I follow Coggshall in attempting to bring acknowledgement of inherent variation back into the analysis of short-*a*.

Chapter 5: Short-*a* Results

5.1 Introduction

This chapter presents data showing the range of short-*a* split systems present in the speech of student members of the CyberPilots of Grumman High School and their mentors. I show that the New York City English (NYCE) complex short-*a* split, which I will continue to refer to as the Tragerian split, is still found in about one third of the speakers, including some young speakers. The majority of speakers aged 14-18, however, have the more widespread nasal split. About half of those students, though, have retained constraints from the Tragerian split in their nasal-split systems.

After having presented so many different approaches to the study of the short-*a* split in Chapter 4, I will first briefly review here the main conditions of the Tragerian split on which I will be basing my analysis. Then I describe my methodology for measuring and analyzing my data in Section 5.3. In Section 5.4, I present the four categories of short-*a* split that my speakers exhibit and how they pattern by the social factors of age, gender, and ethnicity. In Section 5.5, I further discuss variation within the categories presented in the previous section, and also make a case for a single measure to account for both the fronting and raising that is involved in *tensing* short-*a*. In Section 5.6, I explore some additional constraints beyond phonological environment that may affect speakers' pronunciations of short-*a*. Finally, in Section 5.7, I discuss the patterns that emerge in my data, and then how these findings fit in with other recent work such as

Becker's proposed change-in-progress away from a complex split in Manhattan, and the reimagined framing of the NYCE short-*a* system put forth in Coggshall 2017.

5.2 Short-*a* in NYC

5.2.1 Following Environment

As outlined in the previous chapter, the NYCE complex short-*a* split is conditioned primarily by the following phonological environment. The traditionally described Tragerian split puts following front nasals, voiced stops, and voiceless fricatives in the *tense* class and following velar nasals, voiceless stops, and voiced fricatives in the *lax* class. However, there is variation in these classifications, and I follow the model summarized by Coggshall 2017, seen in Figure 27 below.

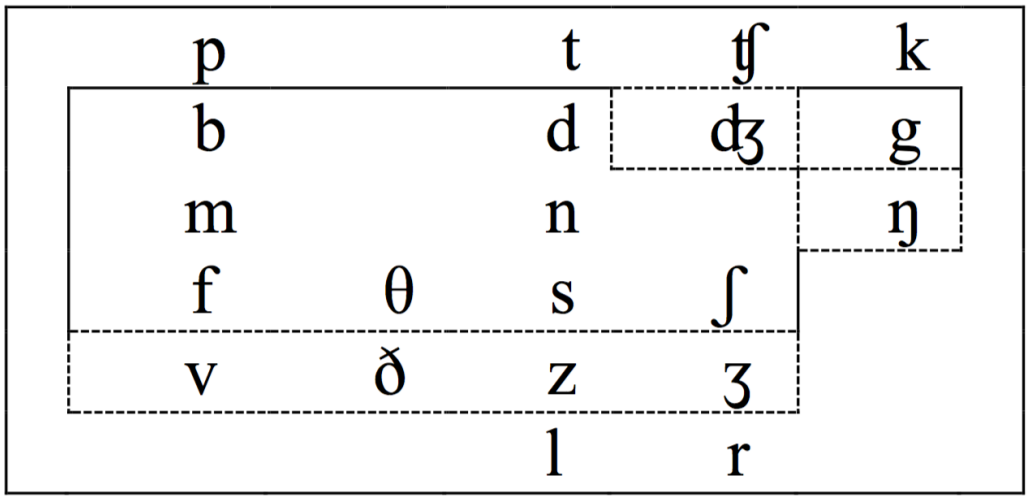


Figure 27: Coggshall's depiction of the following environments that condition the NYC short-*a* split

In Figure 27, the following environments in the inner box are the ones that condition *tense* production of short-*a* (voiced stops and voiceless fricatives). Surrounded by dotted lines are those that have been found to be variable throughout the history of study of this vowel system: voiced fricatives, the voiced affricate, and the velar nasal. In the outer-most box are those environments that condition a *lax* short-*a*: voiceless stops and the voiceless affricate.

5.2.2 Additional Constraints

While conditioning by the following phonological environment is the main determiner of which short-*a* allophone surfaces in the Tragerian split, there are additional patterns and exceptions to be accounted for. I have discussed these in my review of the literature in Chapter 4, and I highlight here the ones I will be focusing on. Four of the constraints are strictly lexical; these constraints pertain to function words (e.g. "had"), abbreviations (e.g. the nickname "Cass," a category perhaps better described as "hypocoristics" based on the scant examples given by Labov when devising this constraint), lexical exceptions (e.g. "avenue"), and learned words (e.g. "carafe"). The remaining three (the Open Syllable Constraint, Inflectional Boundary Closing, and the Initial Condition) are phonologically based patterns. Here I concern myself primarily with the most productive of these phonological patterns: The Open Syllable Constraint, which I describe in more depth below. I also comment at the end of the chapter on the Inflectional Boundary Closing constraint and what my data shows in regard to its persistence in the systems of the speakers in this study.

The Open Syllable Constraint (OSC) holds that short-*a* vowels in an open syllable surface as *lax* where they would be *tense* if the same following segment were tautosyllabic.

For example, the vowel in the word “plan” /plæn/ is *tense,* but the vowel is *lax* in “planet,” where the word is syllabified as /ˈplæ.nət/, with the short-*a* in an open syllable. However, an inflectional boundary causes the short-*a* to act as if it is in a closed syllable, rendering the vowel in, for example, “planning” *tense*, even though it is arguably in an open syllable. This is referred to by Labov 2007 as “Inflectional Boundary Closing.”

It is appropriate for me to mention the Initial Condition and to explain why I am not considering it here. This constraint says that short-*a* will surface as *lax* when it is word-initial even if the following environment would normally condition it as *tense*, for example in the word “aspirin.” While this constraint operates with relative frequency, I have excluded it from consideration because of confounding factors in testing it. Most significantly, the preponderance of word-initial short-*a* tokens in my data occur in open syllables. When this occurs and the vowel is *lax*, it is not possible to determine which constraint is operating.

5.2.3 Underlying Variation

As described in Chapter 4, it is important to note that there is some variation that has always existed in this pattern for the speakers of New York City English. Ignoring this variation and describing the pattern in categorical terms has led to mistaken reasoning and conclusions that the Tragerian split is dissipating, when some of that evidence more accurately just illustrates inherent variation in the system. Coggshall 2017 shows that there was always substantial variation of short-*a* before velar nasals, whereas Becker 2010, for example, used the velar nasal as a primary argument for the system moving away from the Tragerian split in her youngest

speakers. Coggshall also shows significant variation before the intervocalic voiceless fricatives /s/ and /ʃ/, whereas that is usually (following Labov 2007) classed as a categorical *tensing* environment.

In my analysis, I attempt to take into account all of the sites of variation that have been documented in the short-*a* system going back to Trager. Because of this, my threshold for classifying a speaker as having a Tragerian split may seem more permissive than that of previous researchers (Becker and Wong 2010, Becker 2010, Newman 2011, *inter alia).* However, I do so by design, because I believe that phenomena like *tensing* before velar nasals do not mean that the speaker has an "impure" or deficient short-*a* split. Rather, it constitutes a legitimate instantiation of a Tragerian split where this environment has been historically variable.

5.3 Methodology

5.3.1 Measurement and Classification

The data presented here is from the sociolinguistic interviews of 24 students and seven mentors. The interviews were transcribed in ELAN, and then put through the FAVE suite (Rosenfelder et al. 2011) to align the transcriptions to the recordings and to extract the vowel measurements. All short-*a* vowels with a duration of more than fifty milliseconds are measured for F1 and F2 at 30% through the vowel and are normalized across speakers. While the data analyzed here comes primarily from the interviews, some data from the word list and reading tasks are also used as support where noted. Only tokens with primary stress were analyzed, including both syllables in disyllabic compound words (like "snapchat," which came up very

frequently in the interviews). Tokens of short-*a* preceding /r/ (N=38) and /l/ (N=183) were eliminated due to both *r*- and *l*-coloring effects on preceding formant values (Becker 2010:170), as well as the fact that the pronunciation of the vowel in those environments is subject to other regional phonological constraints such as the "Mary-Marry-Merry" merger (Bauman 2013). I also eliminated the word "yeah" (N=2748), as this word is an anomalous use of *tense* short-*a*, with no following environment. Each token in the resulting corpus was coded for following environment, open or closed syllable status, and for the social factors of gender, age, and ethnicity.

The *Atlas of North American English* (ANAE) (Labov, Ash, and Boberg 2006) gives a threshold of 700Hz for F1 to determine whether a short-*a* is *tense* (lower than 700Hz) or *lax* (higher than 700Hz). However, this guideline does not account for the fact that vowels vary with regards to how front or back they are, which is measured by F2. The difference between *tense* and *lax* varies from speaker to speaker enough that the most productive way to classify short-*a* split systems is by plotting each speaker's tokens and visually determining if there are two clear clusters with a significant gap between them, and if so, what description best fits the clusters of tokens (a nasal split or a complex split).

5.3.2 Inclusion/Exclusion of Data

To complete an analysis of each speaker's short-*a* system, I first removed words that would be subject to any of Labov's additional lexical constraints (i.e. function words,

abbreviations, and learned words).[9] In the graphs below I have also not included word-initial instances of the vowel or words subject to inflectional boundary closing. The graphs do include separate following environment categories for open and closed syllables for those classes that would be affected: voiced stops, voiceless fricatives, and front nasals. These are the categories where an open syllable may affect the vowel by conditioning a *lax* variant instead of the *tense* variant that is expected in a closed syllable in a Tragerian system. In speakers with a Tragerian system, we would expect to see these three environments in open syllables pattern with the *lax* categories of voiceless stops and voiced fricatives, and in subjects with a nasal system, we can still test for the Open Syllable Constraint by looking at their short-*a* before front nasals in open syllables. Tokens of short-*a* before velar nasals and voiced fricatives are much rarer than the other following environments, sometimes not showing up at all in a given interview with a speaker. These environments have also been shown to be more variable than the others, as discussed in the previous chapter. The word list data allows me to supplement the interview data where needed, especially in these two categories.

5.3.2.1 Open vs. Closed Syllables

I used the Maximal Onset Principle (Kahn 1976) to decide syllabification, wherein intervocalic consonants are syllabified as the onset of the second syllable to the extent that a licit onset cluster is formed, and any consonant whose addition would create an illicit onset cluster is then assigned to the coda of the preceding syllable. While coding syllables as open or closed, two

[9]The fourth lexical constraint is for "lexical exceptions." Labov's lone example for this constraint is the word "avenue," which I have also excluded.

issues arose where the behavior of short-*a* with regards to the Open Syllable Constraint seemed to be at odds with the syllabification predicted by the Maximal Onset Principle. These two situations were when the short-*a* occurred before a consonant-glide sequence, such as in the words "January," "manufacture," "vacuum," and "canyon." The second situation was when the short-*a* occurred before an /s/-consonant sequence, such as in the words "plastic," "master," "Alaska," and "basket." I briefly discuss each of these observations here.

In the first case, with consonant-glide sequences, a word like "January" would normally (i.e. according to the Maximal Onset Principle) be syllabified as /ʤæn.ju.ɛ.ri/, with the /n/ as the coda of the first syllable and the /j/ as the onset to the second syllable. However, this syllabification would predict a *tense* short-*a* since it would be in a closed syllable followed by a front nasal. Counter to this expectation, words with this pattern tend to be pronounced in New York City English with a *lax* short-*a*, entailing that the Open Syllable Constraint is at work, and the syllabification is actually /ʤæ.nju.ɛ.ri/ with a usually illicit onset cluster of /nj/.

Jaggers 2018 presents evidence that loan words that begin with /Cj/ (such as "Kyoto") are more likely to be adapted by English speakers using hiatus, turning the glide into a full vowel (/ki.o.to/), rather than accepting a /kj/ onset cluster. Loan words with a medial /kj/ cluster (such as "Tokyo") are less likely to undergo the same adaptation, presumably because they can licitly be syllabified with the stop and glide in separate syllables. Generally, the more sonorous a preceding consonant is, the less likely it is to be followed by a glide in a complex onset (Gouskova 2004), so /nj/ would be an even worse onset cluster in English and would more strongly favor hiatus if in initial position.

One possible analysis that accounts for most of my tokens in this category is the case for /ju/ as a monomoraic diphthong (Jaggers 2018). If this is the case, then there is no problem with a /nju/ syllable, as it is a simple onset followed by a vowel, rather than a complex onset. I found the majority of the words in my corpus that ran into this issue did indeed involve /ju/ specifically (e.g. “January”, “manufacture”, “Matthew”, “vacuum”). The word “inaccurate” can be argued to have an underlying /ju/ in the third syllable, even if it is usually reduced in casual speech.

There is one word in my corpus, however, that this theory does not explain, the word “canyon.” It does not contain the /ju/ vowel, but behaves the same way, as if the short-*a* is in an open syllable and /nj/ is the onset to the second syllable. While there is only one word of this type in my corpus, there are others that follow the same pattern (e.g. “lanyard”), and so I cannot dismiss “canyon” as a single lexical exception. This environment has also been noted by other researchers writing about the short-*a* split (Coggshall 2017, Sneller 2018). A second possible explanation for these words is that the element may be for some speakers, a palatal (or alveo-palatal) nasal. For these speakers, this pronunciation would then not be a cluster, but a lone nasal consonant. I coded all such tokens in my corpus as open syllables, based on this phonological evidence, my native intuition, and consultation with other native Long Islanders.

With regard to the second anomalous class of words, the opposite syllabification issue obtained. In a word like “plastic” the Maximal Onset Principle would syllabify the word as /plæ.stɪk/, since /st/ is a perfectly fine syllable onset in English and there is no reason to assign the /s/ to the coda of the first syllable. However, words like this are pronounced with a *tense* short-*a* by New Yorkers, implying that the short-*a* is in a closed syllable environment (i.e. /plæs.tɪk/, followed by a *tensing*-segment, in this case a voiceless fricative). One possible test

would be to see if the /t/ in words like this is aspirated, as it would be syllable initially but not if in a consonant cluster preceded by /s/. However, all words like this in my corpus have the /t/ occurring in an unstressed syllable, which makes the aspiration diminished or completely absent anyway. There is ample evidence in phonology that /s/ is special and behaves in ways not usually allowed by English phonotactics, such as forming tri-consonantal clusters and flouting the sonority hierarchy in consonant clusters. So there is reason to consider that this is just another instance of /s/ acting differently from other consonants in terms of syllabification. I coded all words of this type as closed syllables, based on native speaker consultation and the likelihood that /s/ has exceptional properties with regards to syllabification.

5.3.2.2 Inflectional Boundary Closing

I separated out words in which morpheme boundary closing would be possible. These are words where a short-*a* that is in a closed syllable in the stem of a word is put in a position where the syllable could be opened by the addition of an inflectional morpheme that begins with a vowel (e.g. "-ing"). Recall the example from above: a word like "plan" would have a *tense* short-*a* because it is in a closed syllable and followed by a front nasal. If the "-ing" inflectional ending is added, it would render the short-*a* in an open syllable which would trigger the Open Syllable Constraint and cause short-*a* to become *lax* in the word "planning." However, the Inflectional Boundary Closing constraint predicts that in "planning" the short-*a* actually remains *tense* if the ending is an inflectional ending. Here, I only isolated words where the short-*a* is followed by a *tensing* class of sounds so that the open/closed syllable would actually render a change in whether the vowel is *tense* or *lax*. Words in which the following environment would already

predict a *lax* vowel, and additionally have such an inflectional ending, were not separated from the main data and are included in the base vowel plots below.

I also excluded words that fit the same criteria as above but had derivational endings instead of inflectional ones. I originally identified these through my own intuitions, but then consulted two other native Long Island speakers to find that their intuitions concurred. For example, I found that the word "passage" seems to treated the same way as "passing," even though "-age" is a derivational morpheme and "-ing" is an inflectional morpheme. In Section 5.6.1 below I present evidence that these endings do indeed behave the same way, and that the "Inflectional Boundary" constraint is more accurately a lexical process involving faithfulness to the stem, regardless of the status of the morpheme that is appended.

5.3.2.3 Word Initial Tokens

I separated out all tokens where the short-*a* is word-initial (N=1211) and did not include them in base analyses of following environment. I do not do a specific analysis of the Word Initial Condition here because of the dearth of data left after removing any tokens with confounding factors. There are a large number of words with word-initial short-*a* that also have the vowel in an open syllable, and thus are not useful for isolating the effects of a word-initial constraint separately from an open syllable constraint. This represented 107 of the word initial tokens. Among the words in my corpus that have short-*a* word initially and in an unambiguously closed syllable, there are then a quite large number where the following environment conditions a *lax* pronunciation anyway, and so they also do not help test the constraint (this was the preponderance of the applicable data, N=700). I was then

left with 404 tokens in which the short-*a* is word-initial, in a closed syllable, and followed by a segment that could condition it as *tense* in either the Tragerian or nasal systems. But further depleting the supply of words to test this constraint are the known exceptions to the word-initial constraint already identified by Labov in the formulation of said constraint, where he says that "initial short-*a* with a coda that normally produces *tensing* is *lax* {ask, asterisk}, except for in the most common words {ask, after} (355). "Ask" and "after" were the most common words in this section of my corpus (N=260). This left me with 140 tokens that could be used to test the existence of this constraint. Only 72 of these tokens are followed by nasals, and so could be used to test the constraint in subjects with nasal systems. Previous studies have often failed to account for many of these confounding factors, so I wanted to make sure I was being thorough. However, this thoroughness led to a result where I do not have enough data to adequately test condition across speakers and phonological environments. I conclude my discussion of the Word Initial Condition here, but I suggest that future research on this subject target words in elicitations that could test this condition.

5.4 Individual Speaker Classification

Each vowel plot below shows the values for the pronunciation of short-*a* before front nasals, velar nasals, voiced fricatives, voiceless fricatives, voiced stops, and voiceless stops. By looking at these six categories, it is possible to determine if a subject has a split between *tense* and *lax* variants, and if so, which following environments correspond to which side of that split.

Also incorporated into these graphs are open vs. closed syllable environments, as these prove to be relevant in a majority of speakers and collapsing these categories even in a preliminary analysis would appreciably skew the results by obscuring sites of significant variation. The following environments that condition a *lax* pronunciation of the vowel in any type of short-*a* system (voiceless stops and voiced fricatives) are not broken down into open and closed syllable sub-categories, because that would not be diagnostic of anything. The following velar nasal category is also not broken down. Since this phoneme cannot be the onset of a syllable in English, all vowels in this environment are assumed to be in closed syllables. In addition, this is a class with very few tokens anyway, with many speakers not having any in their interview at all.

I classified my speakers into four types of short-*a* system, each of which I will describe in more detail in the following sections. These categories are 1) Tragerian, 2) Transitional, 3) NYC-Nasal, and 4) Nasal.

5.4.1 Tragerian split

A total of ten speakers (32%) were classified as having a Tragerian system. These include six mentors and four students, all male speakers. The full demographic information for each of these speakers can be found in Table 6.

Subject	Status	Age	Gender	Ethnicity	Religion
Lenny	Mentor	86	Male	White	Jewish
Mr. Spector	Mentor	66	Male	White	Jewish
Mr. Geissler	Mentor	65	Male	White	Catholic
Isaac	Mentor	59	Male	White	Jewish
Mr. Price	Mentor	30	Male	White	Catholic
Tony	Mentor	24	Male	White	Catholic
Daniel	Student	Senior	Male	White	Jewish
Logan	Student	Junior	Male	White	Jewish
Ethan	Student	Sophomore	Male	White	Jewish
Caleb	Student	Freshman	Male	White	Jewish

Table 6: Demographic Data of Speakers with a Tragerian short-*a* Split

An example of a speaker with a robust Tragerian split is Mr. Geissler (one of the head teachers of the club), a white, Catholic male. Figure 28 below shows Mr. Geissler's short-*a* production during his sociolinguistic interview. His means for the three Tragerian *tense* environments (following front nasals, voiceless fricatives, and voiced stops) are clearly distinct from his means for the *lax* environments, and any overlap of the ellipses (representing one standard deviation from the mean) is negligible. The Open Syllable Constraint is also very clear. N-C, S-C, and D-C represent the three Tragerian *tense* environments in closed syllables, while their open syllable counterparts (N-O, S-O, and D-O) all pattern with the other *lax* environments.

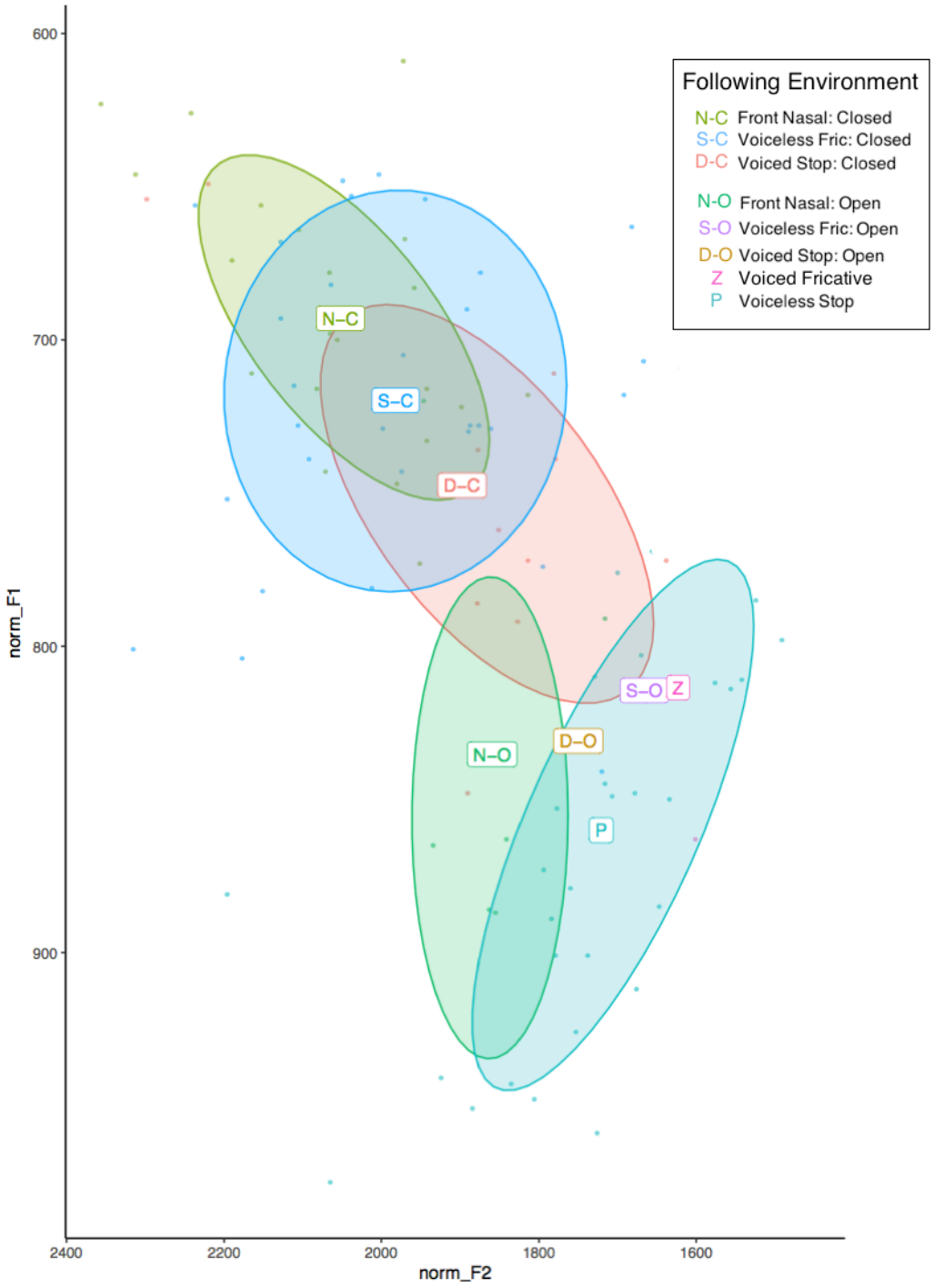


Figure 28: Mr. Geissler's Vowel Plot by following environment

In Figure 28, the means for Nasal-Closed, Voiced Stopped-Closed and Voiceless Fricative-Closed are all on one side of the split (front and high), while the Nasal-Open, Voiced Stopped-Open, Voiceless Fricatives-Open, Voiced Fricative, and Voiceless Stop are on the other side of the split (backer and lower). Mr. Geissler did not have any tokens of short-*a* preceding a velar nasal in his interview. There is variation among the Tragerian speakers in treatment of the velar nasal, and I will discuss further in Section 5.5 how I dealt with it in my classifications.

5.4.2 Transitional System

Four speakers (16%) were classified as having a Transitional system. What distinguishes these speakers from those with a Tragerian split is their pronunciation of short-*a* before voiceless fricatives (the BASH word class). These speakers have their BAD word class on the *tense* side of the split as Tragerian speakers do, but their BASH class is on the *lax* side of the split. All of these speakers also display an Open Syllable Constraint and have velar nasals as *tense*. This category includes one mentor and three students, all of whom are male. Their full demographic information is listed below in Table 7, and Figure 29 illustrates an example of a Transitional speaker's vowels.

Subject	Status	Age	Gender	Race	Religion
Nate	Mentor	31	M	White	Jewish
Paul	Student	Senior	M	White	Jewish
Luke	Student	Junior	M	White	Jewish
Joe	Student	Junior	M	White	Jewish

Table 7: Demographic Data of Subjects with Transitional System

Figure 29: Nate's Vowel Plot by following environment

The distinctive aspect of the Transitional class, the *lax* short-*a* before voiceless fricatives in closed syllables, is very visible on this graph. The large purple ellipsis at the bottom right represents this environment. In Nate's case, his pronunciations for this environment span a wide range (hence the large ellipses), but they are all definitively *lax*. Also note that BASH has lowered farther than even BAT, which is a reoccurring feature of the speakers with Transitional systems and will be illustrated further in the analysis of variation within groups below.

5.4.3 Nasal Systems

Seventeen speakers (55%), all of whom are students, did not have Tragerian or Transitional systems. These students divide into two groups: those that have a full nasal split, and those that have what I am calling an NYC-Nasal split. NYC-Nasal split refers to a system wherein the main phonological conditioning is between nasal vs. oral following environments, but which also displays the Open Syllable Constraint, a feature of the Tragerian system. NYC-Nasal accounts for a little more than half of the Nasal students. This group contains most of the female students in the sample (six out of nine). The full demographic data for NYC-Nasal speakers is listed in Table 8, and the information about full nasal-split speakers (i.e. those *without* the OSC) is listed in Table 9.

Subject	Status	Age	Gender	Ethnicity	Religion
Chrissy	Student	Senior	Female	White	Catholic
Amy	Student	Senior	Female	Chinese	
Carina	Student	Junior	Female	Chinese	
Hannah	Student	Junior	Female	White	Jewish
Leah	Student	Junior	Female	White	Jewish
Chloe	Student	Freshman	Female	White	Jewish
Garrett	Student	Senior	Male	White	Jewish
Alex	Student	Freshman	Male	White	Jewish
Jake	Student	Sophomore	Male	White	Jewish

Table 8: Demographic Data of Subjects with Nasal Split and Open Syllable Constraint (NYC-Nasal)

Subject	Status	Age	Gender	Ethnicity	Religion
Tahani	Student	Sophomore	Female	Indian	Sikh
Diviya	Student	Freshman	Female	Indian	Sikh
Rebecca	Student	Freshman	Female	White	Jewish
Matan	Student	Senior	Male	Middle Eastern	Jewish
Adam	Student	Junior	Male	White	Jewish
Noah	Student	Sophomore	Male	White	Jewish
Tim	Student	Freshman	Male	White	Jewish
Ranjit	Student	Freshman	Male	Indian	Hindu

Table 9: Demographic Data of Subjects with plain Nasal Split (no OSC)

An example of a student with an NYC-Nasal system, with the Open Syllable Constraint, is presented in Figure 30 below.

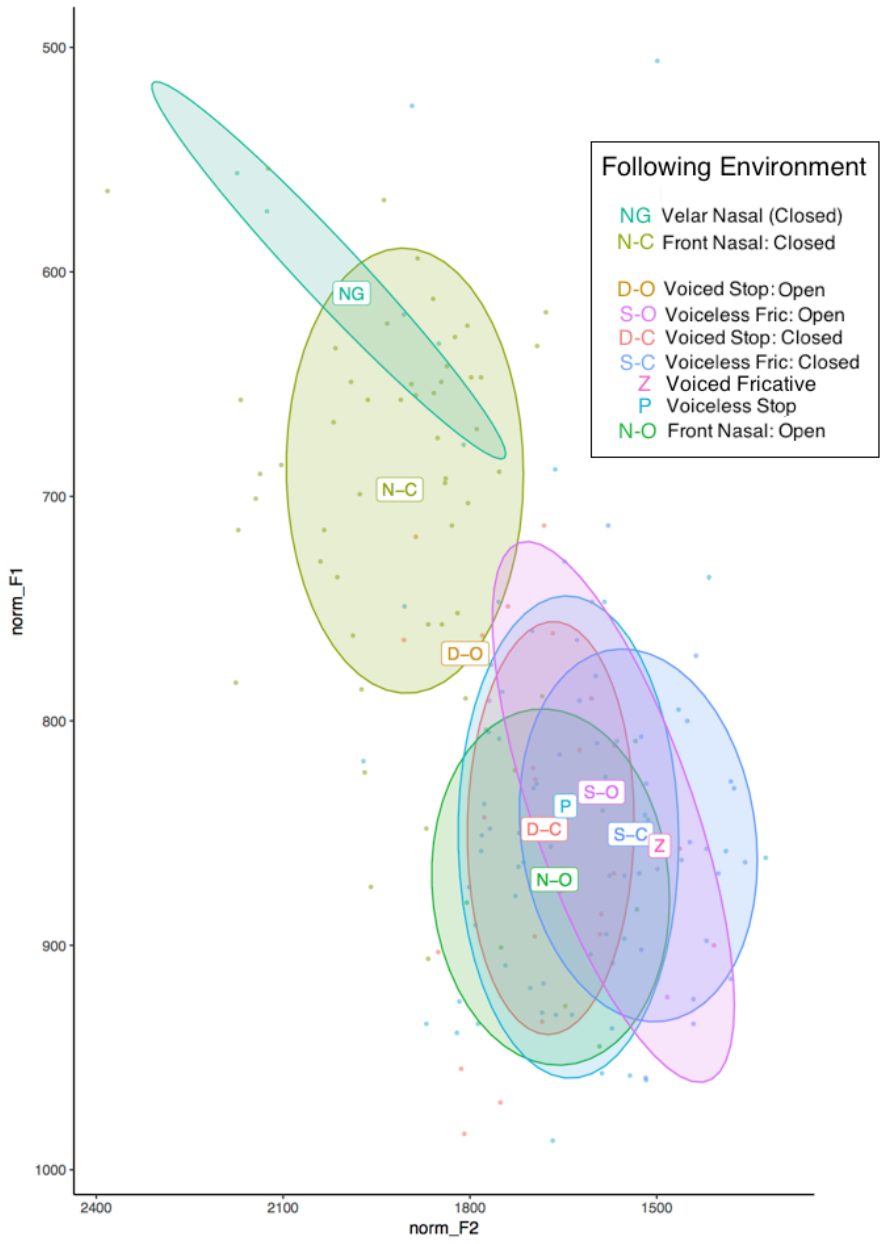


Figure 30: Chloe's Vowel Plot by following environment and open/closed syllable[10]

[10]The environment of following voiced stops and voiceless fricatives in open and closed syllables are plotted separately in this figure, but there is no significant difference in their (all *lax*) pronunciations.

Chloe's plot shows a clear split between the Nasal-Closed and Velar Nasal Categories on one side, and everything else on the other. Crucially here, the category of words with following nasals in open syllables (N-O) patterns with the *lax* categories and its ellipsis does not overlap at all with the Nasal-Closed category. Even though Chloe has a system that is conditioned by nasality, this condition only applies to tautosyllabic following nasals, and the Open Syllable Constraint is still part of her short-*a* split system.

Chloe contrasts against the students who have a full nasal split that is impervious to the open-closed syllable distinction, as shown in Diviya's plot in Figure 31. Diviya is a freshman female of Sikh Indian descent. Her family are recent immigrants to New York (both of her parents were born in India). All three of the Indian students (all of whom have foreign-born parents) have full nasal-split systems. The two Chinese students (both female), one of whom has foreign born parents and one who does not, have NYC-Nasal systems.

Figure 31: Diviya's Vowel Plot by following environment and open/closed syllable

In Diviya's plot, it is clear that the Nasal-Open and Nasal-Closed categories pattern together and are both distinct from the rest of word classes, which are pronounced as *lax*. The velar nasal category is also *tense* (the mean label is located at the top right of the Nasal-Closed ellipsis) but it has no ellipses itself because of how few tokens there are (four or less).

5.5 Variation within Groups

5.5.1 Accounting for both Raising and Fronting

While previous studies do describe *tensing* as a product of both raising and fronting, most of the focus is usually on F1. As mentioned above, ANAE declares a threshold of 700Hz for F1 to call a token *tense* (although the dividing lines drawn on the vowel plots throughout ANAE are often at different exact locations). Those studies that *do* take into account F2, perform separate analyses of each formant in tandem. Here I attempt to develop one measure that can represent both the raising and fronting aspects of *tensing* in one number, so that future statistical analysis might be able to move away from separate analyses of F1 and F2 that do not show the whole picture about in one view. Towards this end, I introduce in this section a *Tensing Score* that can be simply calculated and used, to some extent, to combine the two components of tensing into one quantifiable measure.

For this *Tensing Score*, I draw on Labov, Rosenfelder and Fruehwald 2013 (LRF) and their analysis of one hundred years of sound change in Philadelphia. As shown in Figure 32, LRF argue that the best measure to approximate movement up and down the front periphery of the

vowel space is to calculate F2-(2*F1). This calculation allows a higher F2 value (corresponding to a fronter vowel) to set a higher score initially, and then the lower the F1 value is (the higher the vowel is) the less that will be subtracted, yielding a higher score. Back and low vowels will therefore have low scores, while front and high vowels will have high scores. The range of this score is about -500 to 1500 if looking at the whole vowel space, but for short-*a*, the values are mostly between -300 to 900.

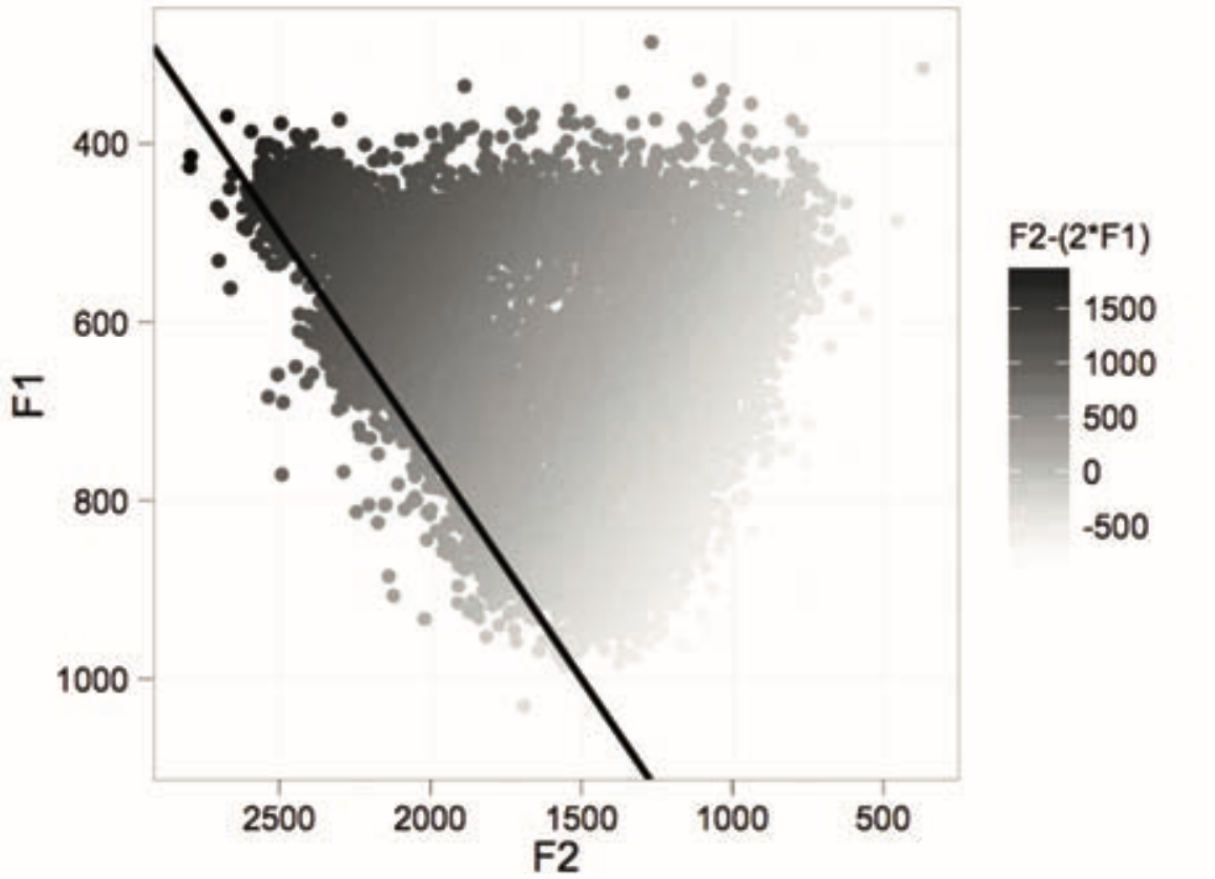


Figure 32: Measurement of location along the front diagonal of the vowel space as derived by F2 - (2 * F1), from Labov, Rosenfelder, and Fruehwald (2013:40)

Figure 32 contains data from the entire vowel space. Tokens are darker if they have a higher score (are front and high) and are lighter the lower the score gets. The diagonal line superimposed on the graph is the front barrier of the vowel space. You can see that at the top of the line, the tokens are all darkest black and get lighter as they go down the diagonal line. This shows the correlation between this number and movement along the front of the vowel space. I use this number as a *Tensing Score*, to serve as a single numeric value that takes into account both dimensions of short-*a tensing* or *laxing*. While it would be equally possible to assign *Tensing Scores* to vowels in open syllables, in the present discussion I limit myself to closed syllables for simplicity in illustrating this new method.

5.5.2 Key to reading the figures in the following sections

In the figures below, I attempt to combine a great deal of information into a single table in order to make an illustration of variation within each short-*a* system digestible. I use color to incorporate an additional dimension into the table, and I use the *Tensing Score* as the measure rather than values of F1 and/or F2.

In the tables below, the colors each represent a word class (all are closed syllables). The shades of orange represent classes that are *tense* in the classic Tragerian split (BAN, BAD, and BASH), and the shades of green represent the classes that are *lax* in the classic Tragerian split (BANG, JAZZ, BAT). Each speaker heads a column in which these six classes are ranked by highest Tensing Score to lowest Tensing Score for that speaker. A double line is placed where the

speaker's short-*a* split occurs.[11] For example, if a speaker has a completely traditional Tragerian split, all of the orange boxes would be above the double line and all of the green boxes would be below the double line, as illustrated in Figure 33.

Figure 33: Color key for the figures in the following sections. In this example, the colors are ordered from highest Tensing Score to lowest, as they would appear in a classic Tragerian split.

Something that can be seen frequently is the lightest of the green shades (BANG) appearing above the split line for speakers with Tragerian or Transitional systems. This represents the velar nasal class being pronounced as *tense* rather than *lax*, a category which was focused on in Chapter 4 as the main feature that many have pointed to as evidence of a change in progress, but which Coggshall shows has always been a variable category. Because of this, as well as the paucity of tokens for the velar nasal in general, I do not disqualify a speaker from being Tragerian or Traditional for having a high Tensing Score for the BANG class.

[11]While I believe the Tensing Score is illustrative in important ways, it does not allow for as clear of a visual representation of where a given speaker's split is located. This still is most productively done by studying vowel plots. The double lines indicate where I have determined each speaker to have the most prominent split in their short-*a* production, having looked at the data in a number of different configurations.

Another pattern to be familiar with is what a Nasal system looks like in this model. In a Nasal system, only two boxes will be above the split line, BAN and BANG, as is illustrated in Figure 34.

Figure 34: Example of what a nasal split looks like, with the darkest orange (BAN) and the lightest green (BANG) above the double line that represents the speaker's split.

It should be noted here that I do include a label at the bottom of the figures indicating which students have NYC-Nasal splits and which have full nasal splits, but that differentiation does not bear on the information in these graphics, as the qualifying characteristic of NYC-Nasal speakers is having an Open Syllable Constraint, and only closed syllable are examined here.

5.5.3 Mentors

Figure 35 shows a ranking of following environments in closed syllables for each mentor. They are ordered from most *tense* at the top to most *lax* at the bottom. The number within each box is the speaker's average *Tensing Score* for this phonological environment. The box is left blank if the speaker had no tokens for that environment in their interview. The double line in each column delineates the split for each speaker, with their *tense* classes above the line and *lax*

classes below it. The number above each mentor is his age. All of the mentors are white and male; Mr. Price, Mr. Geissler and Tony are Catholic, while the remaining four mentors are Jewish. Religion does not readily correspond to any trends, so it is not included in the chart.

Mentors						
60	86	61	30	24	59	31
Mr. S	Lenny	Mr. G	Mr. P	Isaac	Tony	Nate
774	858	666	684	630	646	569
618	620	542	625	366	334	489
541	500	403	505	343	319	461
118	275	1	251	309	276	288
-24	230	-4	81	99	150	23
	144		-227		49	
Tragerian						Trans'l

Figure 35: Tensing scores for each following phonological category in closed syllables for each of the Mentors, ranked from *tensest* to *laxest*.

Six out of seven of the mentors have a Tragerian system. The first apparent point of variation within this group is that two of the mentors, Tony and Isaac, have BANG on the *tense* side of their split (above the double line), while the rest have it on the *lax* side.[12] Mr. Geissler did not have any tokens of BANG in his interview, so that cell is blank, but his word list results showed that he is invariably *lax* in this environment. Despite the variation in the treatment of velar nasals, what ties these speakers together as one group (Tragerian) is that crucially their BAD and BASH classes are *tense*. These are the two main classes diagnostic of a Tragerian split, because BAN is *tense* in all systems, while BAD and BASH are only *tense* in Tragerian systems.

[12]In contrast to his interview data, Isaac's word list data showed all *tense* tokens of the BANG class. This is a point that needs further investigation.

The order in which each speaker has their *tense* classes ranked is variable, but BAN is usually the most *tense*. Mr. Spector is the only mentor who *tenses* BAD more than BAN.

The highest Tensing Scores throughout all categories are seen in Lenny, who at 86 years old is the oldest participant in the study. Regardless of the split of the categories, Lenny's short-*a* is more *tense* overall. Mr. Spector has the largest range of Tensing Scores, spanning 911 points. His most *tense* class is BAN with a score of 684, and his least *tense* class is JAZZ with a score of -227. This is by far the lowest score, or *laxest* pronunciation, in any category by any of the mentors. Most of the mentors do not drop into negative numbers at all (but it will be shown below that negative numbers are prevalent among the nasal speakers). It should be noted that JAZZ, the voiced fricative class, has a very low token count across the board, so this extremely *lax* score for Mr. Spector might be based on a paucity of tokens.

The only mentor with a Transitional system is Nate. His BAD class is clearly *tense*, indicating that he does not have a nasal system and is adhering to the strongest indicator of a Tragerian system, but he deviates from the Tragerian system in his BASH category. Nate's production of short-*a* before voiceless fricatives is extremely *lax*, even more so than his voiceless stop tokens. This is a trend that reoccurs within the group of students who have Transitional systems. BAD is *tense*, so it is not a nasal system, but BASH seems to emerge as the class most likely to become *lax* when transitioning away from the Tragerian system.

5.5.4 Students

5.5.4.1 Male Students

I will first present the white male students (who make up the majority of the sample) and then I will present the female and Asian students in the next section. In Figure 36, I present the male students who have Tragerian or Transitional systems, and in Figure 37 are the males that have NYC-Nasal or nasal-split systems.

White Male Students						
Caleb	Logan	Ethan	Daniel	Luke	Joe	Paul
685	614	820	853	382	427	893
518	571	438	540	297	402	334
394	439	411	495	288	309	326
369	392	361	363	121	195	244
109	181	86	245	88	176	209
30		77		-10	128	45
Tragerian				Transitional		

Figure 36: White Male Students with Tragerian or Transitional Short-*a* Systems

There are four students who have the Tragerian split: Caleb, Logan, Ethan, and Daniel. All of them have the velar nasal (lightest green) category above their split line in the *tense* category, as did Isaac and Tony, two of the mentors with Tragerian splits. There are a few reasons why I have chosen not to disqualify a speaker from being classified as Tragerian if their velar nasals are *tense*. First, velar nasal is the most prominent feature that I highlighted in Chapter 4 in my discussion of the importance of bringing underlying variation back into the analysis. While a *tense* velar nasal has been used as an indicator of a change away from a

Tragerian system, I believe that this may be misguided, as it ignores the inherent variation. That said, I recognize that there is a difference between intraspeaker variation of short-*a* in this environment, and a speaker who consistently produces short-*a* before velar nasals as *tense.* On a practical level, tokens of this condition are extremely uncommon, and many speakers do not produce any at all during their interviews. This makes it difficult to do any analysis of intraspeaker variation, since most speakers have one or zero tokens. On a more theoretical level, if a speaker does treat velar nasals as invariably *tense*, they are not introducing a new condition, but rather they are taking something that is already present variably and making it invariant. They are not changing what is possible in the Tragerian system, so, while this may be a change, I do not consider it one that disqualifies them from being categorized as Tragerian.

Figure 36 also shows the three white male students who have Transitional systems: Luke, Joe, and Paul. As mentioned above, this category is defined by having *lax* BASH, which is visually represented in the above figure by the shade of orange that is under the split line for these three speakers. *Lax* BASH is different from *tense* BANG in that it does not have a history of being observed as variable, and it consistently has more robust quantities of data to analyze. It is also of note here to see that Transitional speakers not only have *lax* BASH, but in three out of four of them (Luke, Paul, and the mentor Nate) BASH is radically lowered, below BAT, showing double digit and negative Tensing Scores.

Figure 37 shows the Nasal counterparts of the white male students in the above figure. As a reminder, while I have noted which speakers have a nasal split and which have an NYC-Nasal split in the chart, that does not bear on the data presented here, which only includes closed syllables.

White Male Students						
Garrett	Alex	Jake	Adam	Tim	Noah	Matan
426	562	865	563	655	510	563
358	471	353			436	
121	147	282	358	216	220	93
79	82	234	223	167	194	43
51	-10	176	184	-41	95	15
34	-71	142	49	-106	17	-137
NYC-Nasal			Nasal Split			

Figure 37: White Male Students with Nasal Short-*a* Systems

Figure 37 shows the seven white male students who have Nasal systems. Adam, Tim, and Matan did not have any tokens of the velar nasal in their interview, but based on their word list data, I was able to confirm that they do pronounce short-*a* as *tense* before velar nasals. I represented that visually here by putting the light green box above the split line, even though it does not contain a value for Tensing Score.

The four speakers who do have tokens of the velar nasal (Garrett, Alex, Jake, and Noah) all rank BANG *tenser* than BAN. Under the line, in their *lax* categories, there is no consensus on which order the four categories appear. It should be observed, though, that five out of the seven show the same super-*lax* BASH as the Transitional speakers did. The BAT word class, the following voiceless stop condition, has often been considered the *laxest* category, and used to bound a speaker's *tense/lax* vowel space (especially in the description of continuous systems), so it is notable that so many of these speakers are pronouncing BASH as their *laxest* environment rather than BAT.

5.5.4.2 Female and Asian Students

Figure 38 presents the five white female students in the study. All five have a Nasal system rather than a complex system. While four out of five of them do have the Open Syllable Constraint (and so are NYC-Nasal) there are notably no females with a Tragerian or Transitional short-*a* system.

Looking at the relative rankings of the word classes we can see that here too, the speakers who have tokens of BANG pronounce them as more *tense* than BAN. Different than their male counterparts, though, is that BASH is not the *laxest* category for any of them. The magnitude of the Tensing Scores are also overall larger for the white female students than for the white male students. Most of the Tensing Scores for nasals for the males are under 600, while the females have several that are in the 700s and even 900s. Becky has the highest Tensing Score in the study, with 998 for her velar nasals. Another generalization that can be made is that the female students have larger gaps between the scores above and below the split line than do the males, indicating a more substantial gap between their *tense* and *lax* pronunciations.

White Female Students				
Chrissy	Chloe	Leah	Hannah	Becky
403	771	592	956	998
	519	537	559	728
191	-16	180	-45	139
135	-29	148	-132	48
96	-159	123	-263	-17
-3	-216	-26	-312	-31
NYC-Nasal				Nasal Split

Figure 38: White Female Students with Nasal Short-*a* Systems

Finally, Figure 39 shows the Tensing Scores for the Chinese and Indian students in the study. They all have Nasal systems (although both Carina and Amy are NYC-Nasal, while Tahani, Diviya and Ranjit have full nasal splits).

Chinese Students		**Indian Students**		
Carina	Amy	Tahani	Diviya	Ranjit
498	658	721	524	527
123	550	328	424	
79	20	15	67	253
-71	-3	6	-36	148
-146	-9	-155	-137	112
-235	-256		-260	-32
NYC-Nasal		Nasal Split		

Figure 39: Chinese and Indian Students with Nasal Short-*a* Systems

One difference that is immediately apparent about this group compared to the previous ones is that for those who have tokens of both front and velar nasals, only one out of the four (Tahani) has the velar nasal as the *tenser* of the two, whereas all of the white students with nasal systems had BANG ranked *tenser* than BAN. Like the white male students (but not the white female students) all the students in this chart have BASH as their *laxest* category rather than BAT.

This section has allowed us to examine some of the variation present within the four major categories of short-*a* system in a way that, I hope, was a bit more visually accessible than comparing full plots of all of each speaker's vowels. While there are, of course, advantages to looking at all the data points, this is particularly cumbersome for analyses of short-*a.* In bringing the representation of the data down to one dimension, the Tensing Score, and using colors,

patterns can be noticed within groups more easily, while still taking into account both raising and fronting as components of *tensing*.

5.6 Two Short Notes

5.6.1 Morpheme Boundary Closing

The literature reviewed in Chapter 4 often presents the Open Syllable Constraint and the Inflectional Boundary Closing constraint as if they were two separate, equal constraints. But Inflectional Boundary Closing is actually a corollary, or an exception to, the Open Syllable Constraint. The OSC causes a token to become *lax* in an open syllable when it would have been *tense* due to its following environment, EXCEPT for when the reason the syllable is rendered open is the addition of an inflectional ending.

While analyzing my data, I noticed that this did not only seem to apply to inflectional morphemes, but derivational ones as well. I have not found a source in the literature for when or why this constraint has been referred to as exclusively "inflectional" boundary closing, but I propose that it is actually a faithfulness constraint to the root of the word, blind to the nature of the morpheme suffix that is added. If the stem of the word is closed and has a *tensing* following environment, then short-*a* is rendered *tense*. When a morphological ending is then added, it does not cause the stem to change, overriding the Open Syllable Constraint, which only applies to open syllables that do not span morpheme boundaries.

In my data, I found that words like "graphic" and "planner" have *tense* short-*a*, even though their morphological endings (-ic, -er) are derivational. It is very difficult to test this

statistically because it is a very specific condition, and the number of tokens that could be used to test it are extremely small. In Figure 40, I make an attempt to represent my assertion graphically.

This plot contains twenty tokens from three speakers; fifteen are inflectional and five are derivational. These are the only tokens by speakers with the Open Syllable Constraint that have a following front nasal. If I added the other Tragerian *tense* environments, I would have to limit the speakers to only those with a Tragerian system, and that would further diminish the data available. Using only following front nasals allows me to include the data from Tragerian, Transitional, and NYC-Nasal speakers.

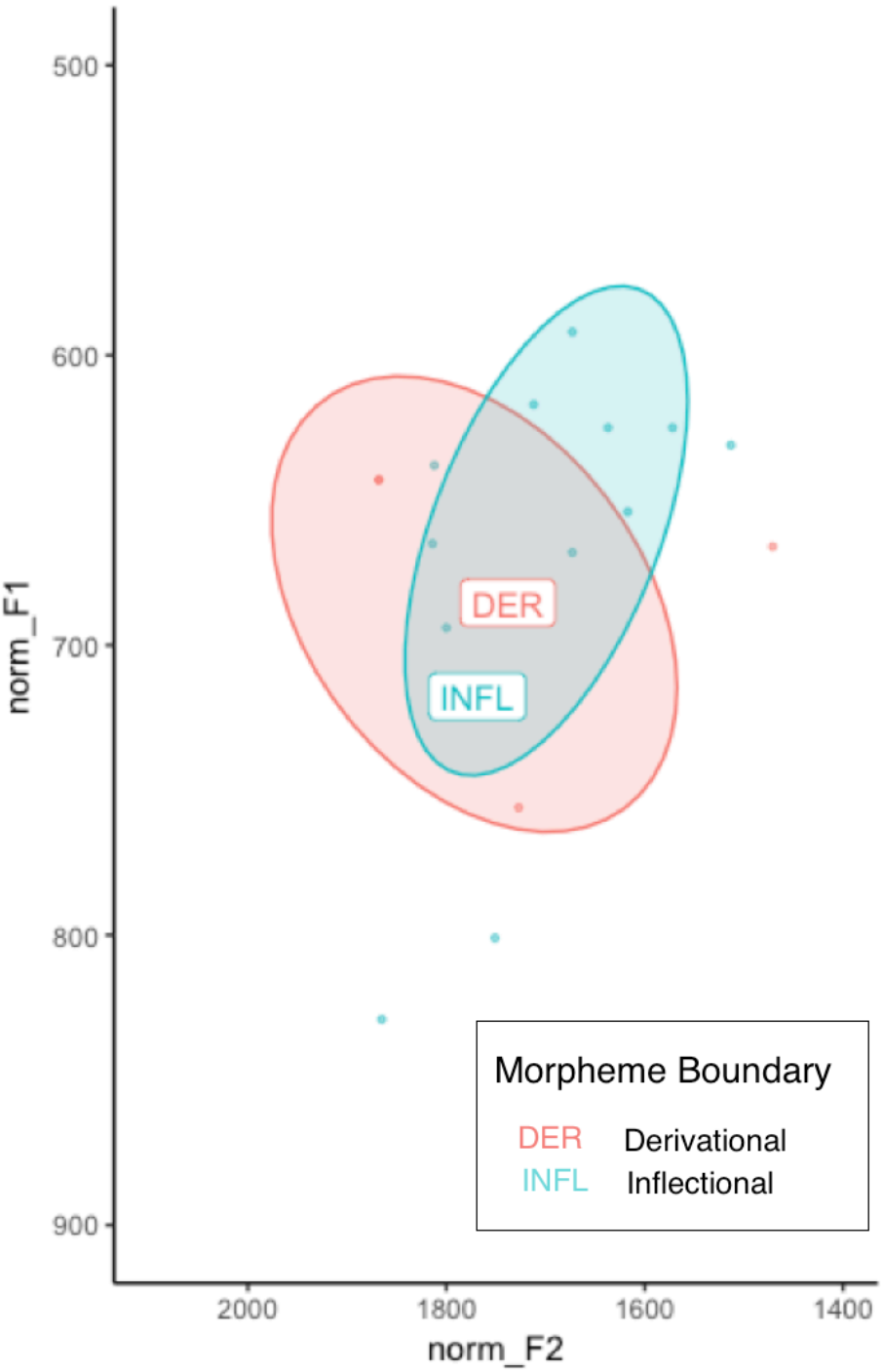


Figure 40: Tokens of short-*a* produced by speakers with the Open Syllable Constraint, illustrating that inflectional and derivational boundaries are treated the same in terms of syllable closing.

What this graph shows it that in the few cases where I can compare, there is no appreciable difference in how the two types of morphemes are treated. Their mean F1 and F2 are very close together. There is not a split between inflectional endings and derivational endings. Granted, this is a small amount of data, but I also have support for this assertion from consultation with native speakers. Future research could test this further by including specifically contrastable words in reading passages or word lists.

5.6.2 The Minimal Pair that Shouldn't Be: The Case of "Can"

While I did not analyze function words in this dissertation, I did pay special attention to the one prominent "minimal pair" that seems to exist between the two variants of short-*a* for NYCE speakers. This is noteworthy because of the persistence of this minimal pair even in speakers who have a complete nasal split and *should* pronounce both with a *tense* short-*a*. My word list task contained the items "I can" and "tin can" to prompt the subjects to produce the word as they would in each context, and almost every single subject had a contrast between the two. Figure 41 illustrates a stark example of this phenomenon. Tahani (Sophomore, Indian) has one of the most distinct nasal splits out of all the students in the study. Intriguingly, as can be seen in this vowel plot, she maintains a very distinct difference between the auxiliary word ("I can") and the content word ("tin can").

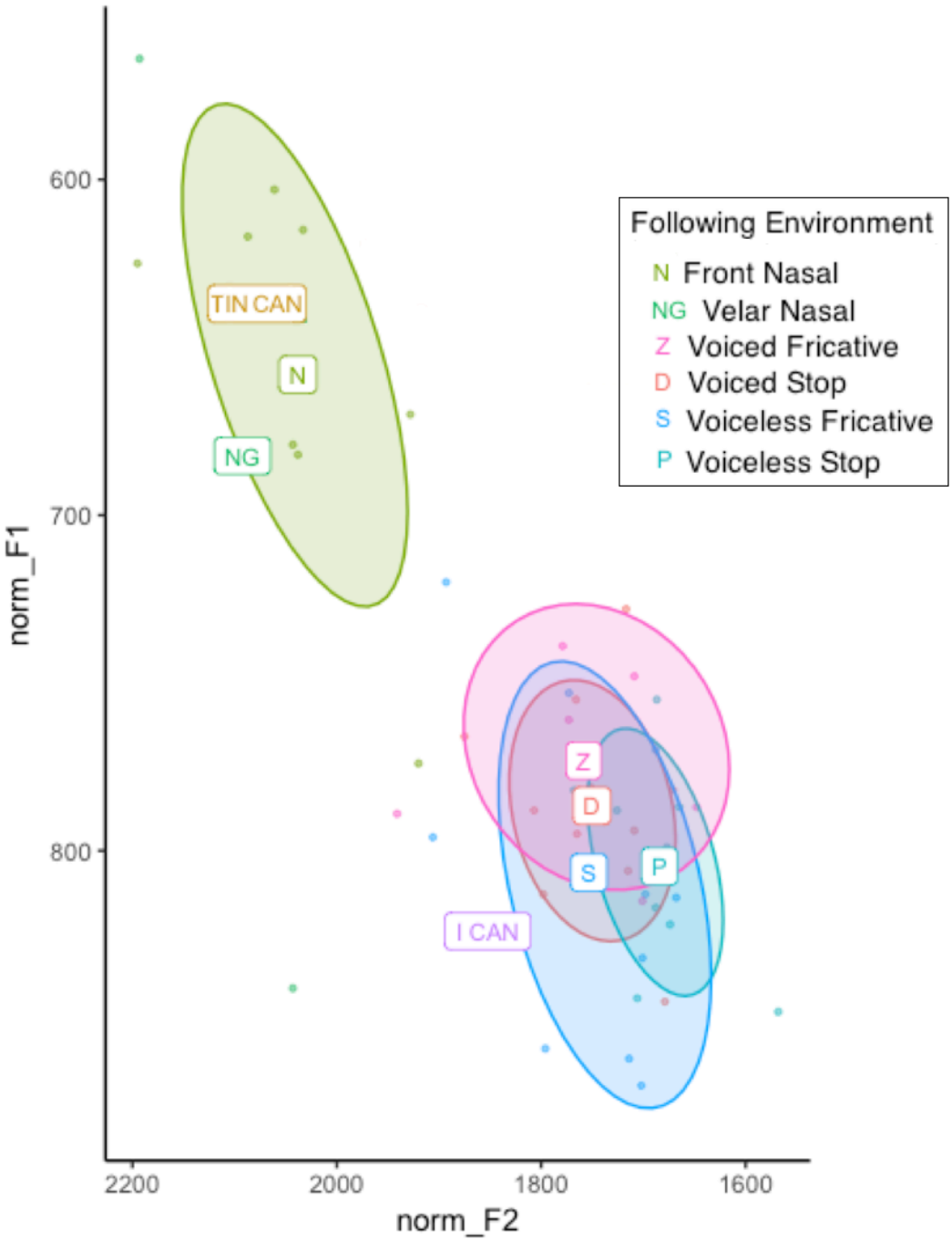


Figure 41: Vowel plot of Tahani's Word List data, showing Nasal Split with the CAN-CAN minimal pair

Tahani has a clear gap between nasals (front and velar) and all other environments. She does not have any difference between open and closed syllables, so those distinctions are omitted from this plot. However, in Figure 41 we can see that she still maintains a sizable distinction between the short-*a* in “I can” and “tin can.” She uses a *lax* pronunciation when saying “I can,” but a *tense* one if talking about the noun.

This is not an anomaly unique to Tahani; it is an overall trend in the sample that this minimal pair has persisted, even for the students with a Nasal or NYC-Nasal system, for which these two words would be expected to be homophones. In Figure 42 below I have plotted the mean values of “I can” and “tin can” from the word list data of all speakers, grouped by short-*a* system category.

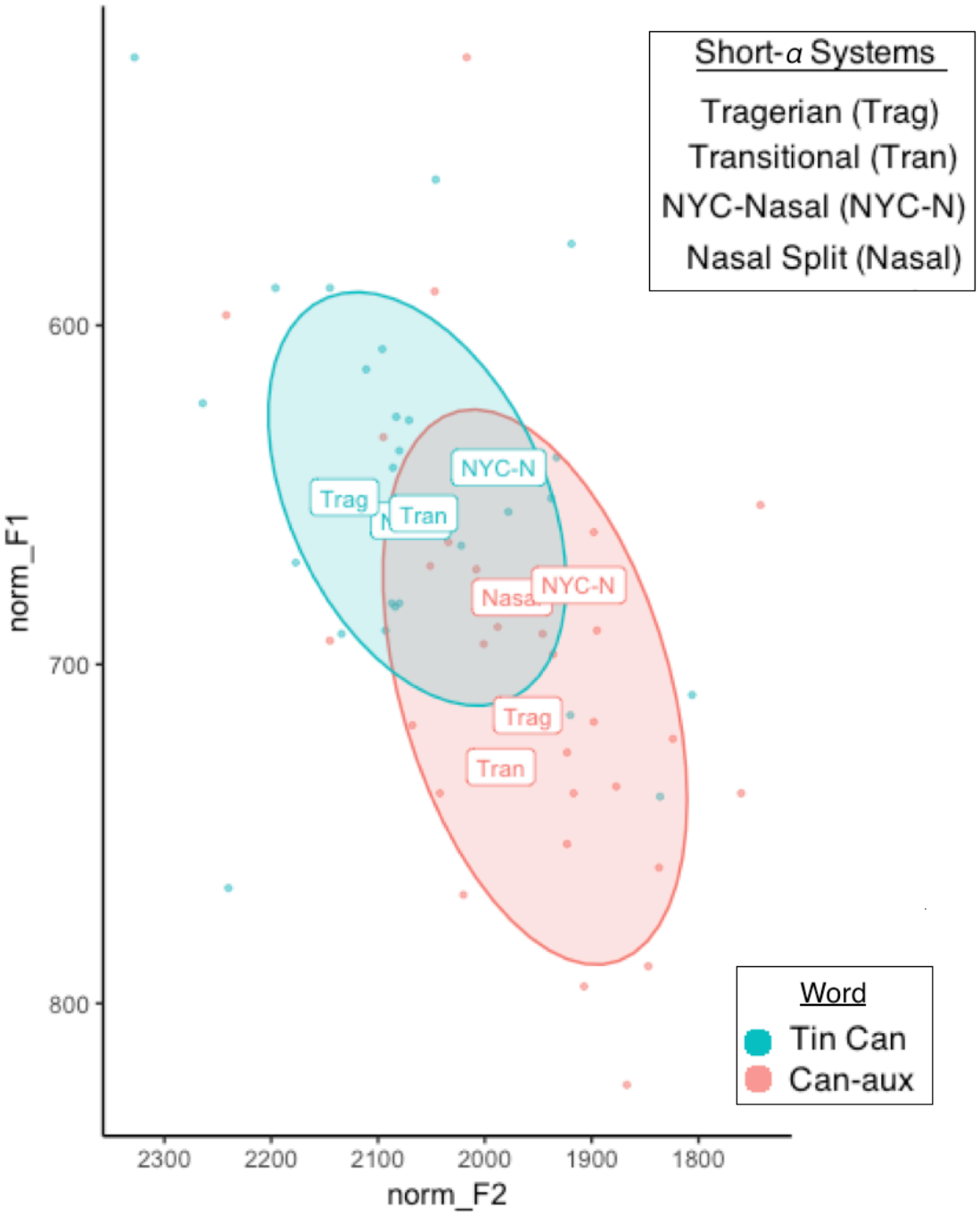


Figure 42: Mean values for the two CAN words across each short-*a* system group[13]

[13]The overlap of blue labels reads "Nasal" underneath the visible "Tran" (Transitional).

All four groups have their CAN-aux (pink) *laxer* than their CAN-noun (blue). For the two Nasal systems, both pronunciations are above 700Hz, so they are both pronounced somewhat *tense,* but "I can" is still more *lax* than "tin can." The placement of these means also obscures some variation within the Nasal system speakers; as seen above, Tahani has a drastically *laxed* "I can," but that is lost in this graph because of using only means.

5.7 Discussion

While only qualitatively analyzed here, there are several patterns that emerge regarding social factors correlating with short-*a* pronunciation. The first that I will discuss is gender. All nine female students have a Nasal system, none of them have Transitional or Tragerian systems. If we take the move towards a nasal system as a change in progress from below, as suggested by Becker 2010, then this trend follows Labov's Principle 3, that in a change from below the level of consciousness, females are more likely to use the innovative form. However, of those nine female students, six of them are NYC-Nasal rather than full nasal split. This means that while they have the more innovative nasal system, it is the female students who are more likely to have conserved an element of the traditional complex split, the Open Syllable Constraint.

There are also patterns to be seen in the data in regard to ethnicity. The three students of Indian descent all have the nasal split. All of these students have parents born in India and represent the most recent immigrant families to this community. They also have had exposure to languages other than English in their homes. It is very noteworthy that Tahani especially, produces "I can" as *lax* as she does. The only possible explanation I can put forward for this, is

that “I can” is such a common and basic phrase, that it would stick out to other children who do have the NYC complex split, and perhaps just that one specific lexical item could have reached the level of consciousness. I can imagine a child on Long Island being mildly chastised or made fun of by their peers for saying “I can” with a *tense* short-*a*. This is only a guess, of course, but it is a very noteworthy fact that this minimal pair has persisted.

Contrasting with the Indian students are the Chinese students. One of them, Amy, has parents who were born in China, but Carina’s parents were born in the US. This could somewhat explain why Carina does not have a full nasal split, as the only one of this group without foreign born parents, but I do not have an explanation for Amy, who has a similar immigrant history to the Indian students.

Finally, age also presents itself as a relevant factor in this analysis. The mentors all fall into the complex split categories, either Tragerian or Transitional, while the majority of the students have one of the nasal systems. This supports a change in which younger speakers are moving away from the Tragerian split and towards a full nasal split. There are some students, however, who do have Tragerian systems. This would not be expected if the change that Becker 2010 proposed was progressing at the same rate in Nassau County as she found it to be in Manhattan. In Chapter 7, I discuss more in-depth a couple of the students who emerge in this chapter and the next as having the most conserved NYCE features despite their young age and offer some more details about their family histories and local identities that may explain these results.

The data presented in this chapter has explored the pronunciation of short-*a* by the 24 students and seven mentors in this study. The traditional Tragerian short-*a* split has many

conditioning factors, and I tried to take into account as many as I could while still being able to illustrate overarching trends. I took a middle ground between the approaches of previous work that entirely relied on only mean F1 to gauge *tensing*, and the work of researchers like Cohen and Coggshall that painstakingly considered each possible conditioning factor separately, allowing an extremely deep but narrow analysis. Researchers employing both of these approaches have greatly influenced my work and have given me a bedrock on which to build. I hope that I have added to the understanding of this very complex phonological system.

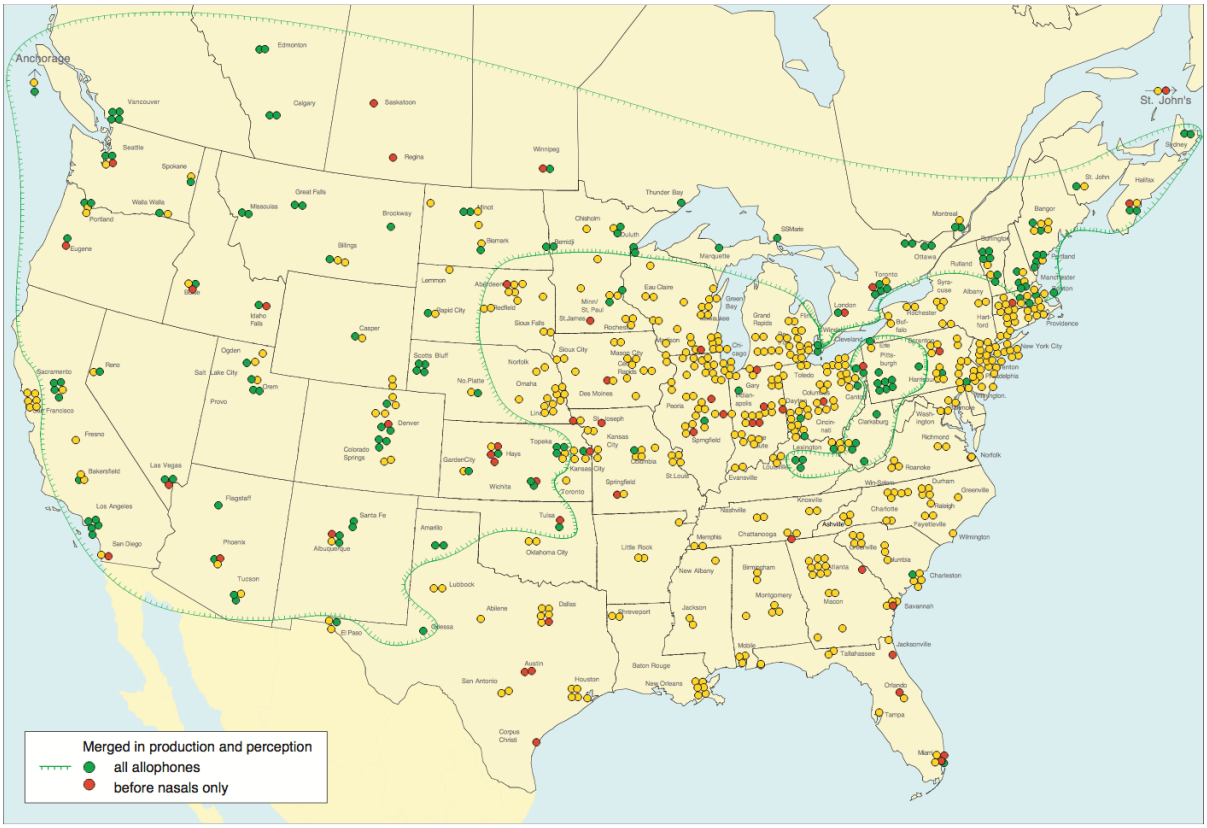
Chapter 6: Raised-THOUGHT and the THOUGHT/LOT Distinction

6.1 Introduction

This chapter presents an analysis of the realization of one of the most salient phonological features of NYCE: the low back distinction between THOUGHT /ɔ/ and LOT /ɑ/, with particular focus on the raising of the THOUGHT-vowel.

6.1.1 The Low Back Vowels

In the mid-Atlantic states, including New York, THOUGHT and LOT are distinct vowels and do not participate in the merger of the low back vowels that is seen in much of the rest of the country (Labov, Ash, & Boberg 2006). The map below from the Atlas of North American English uses yellow dots to represent areas where the two vowel classes are distinct and green dots to represent where there is a complete low back merger. Red dots represent a partial merger, phonologically conditioned by a following nasal.



Map 9.1. The low back merger

Figure 43: Representation of the status of the low back merger across the United States, from the Atlas of North American English (2006:60)

As seem on this map, these two vowels are merged into /ɑ/ for much of North America, including Canada, the western U.S., the north of New England, and western Pennsylvania (Labov et al. 2006). A phonemic distinction between /ɑ/ and /ɔ/ remains in the South, the Inland North, and the Mid-Atlantic region, where a dense cluster of yellow dots is visible on the map.

An apparent-time decrease in the phonemic distance between THOUGHT and LOT has been reported in multiple recent studies of New York City English (Becker & Wong 2010, Wong 2012, Wong & Hall-Lew 2014, Wong 2015). While this increasing degree of overlap between

the two phonemes could raise the possibility of the low back merger appearing in this region, this does not seem to be the case (but see Dinkin 2011 about a possible encroachment of the merger into upstate New York). Evidence of the stability of the phonemic distinction in Nassau County, even among the youngest speakers, will be presented in section 6.3 below.

6.1.2 THOUGHT-Raising

Besides resisting the low back vowel merger, the New York City dialect is also known for a pronunciation of the THOUGHT-vowel that involves raising of the nucleus and also sometimes a diphthongal quality characterized by gliding. Most research on the variable focuses on the height (F1) of the vowel, and this is the approach that I adopt in my analysis as well. Figure 44 below shows a simplified diagram of the American English vowel space with the position of raised-THOUGHT added. The raised variant is above the line marked 700Hz[14] and also farther back in the F2 dimension.

[14]It is useful to remind the reader here that the value of the first formant (F1) corresponds inversely to the height of a vowel. Higher vowels have smaller F1 values and low vowels have larger F1 values. So "above the line" in this figure corresponds to values *below* 700Hz numerically.

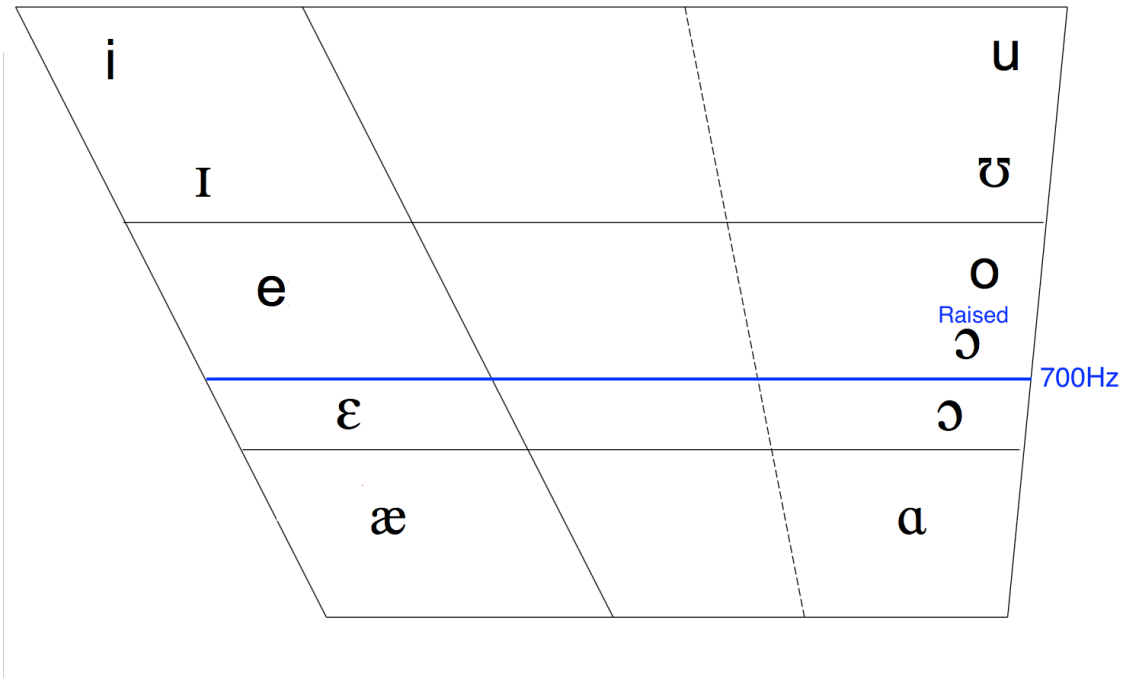
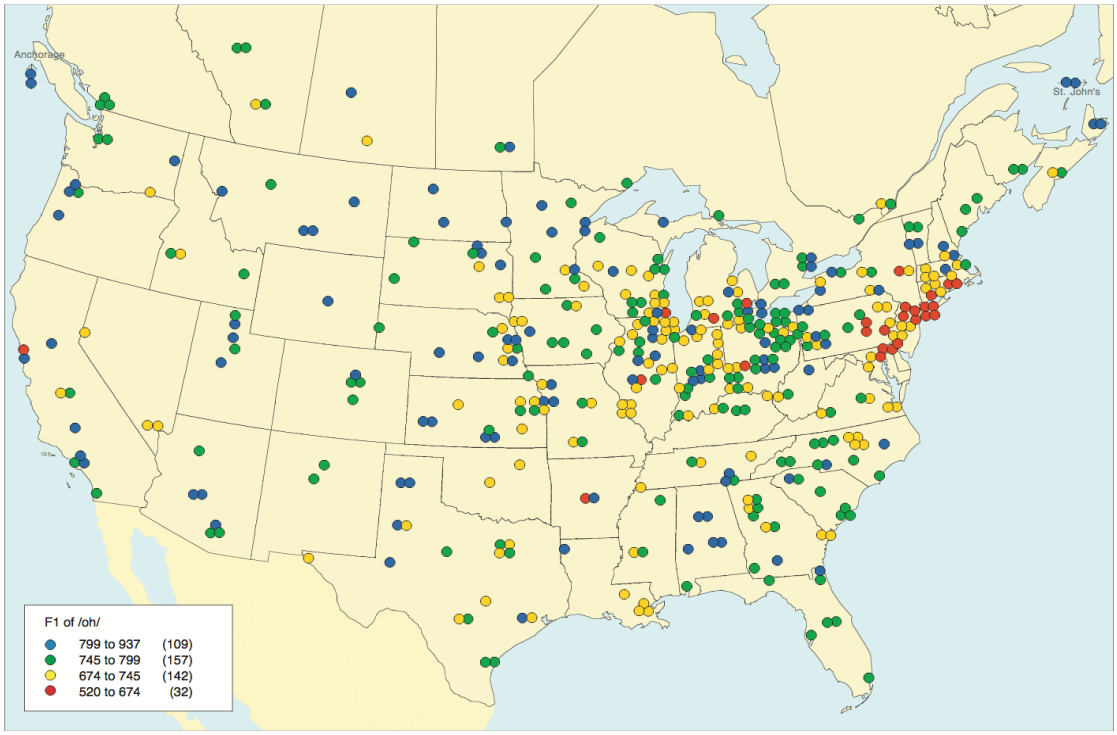


Figure 44: Abstract representation of English vowel space with the location of NYCE raised THOUGHT added.

The line in the above figure is drawn at 700Hz because this is the threshold that ANAE establishes for a token of THOUGHT to be considered raised,[15] and the convention currently used by most researchers. I will also use this number as a benchmark in the remainder of this chapter. The map in Figure 45, from ANAE (108), shows ranges of F1 values for THOUGHT across North America. The lowest F1 values (between 520 and 674Hz) are marked in red, and distinctly cluster around New York City, as well as nearby areas in the mid-Atlantic states.

[15]700Hz is the number explicitly stated in ANAE, but in many of the charts presented therein, including the map presented here, the cutoff is actually usually a lower number. Here it is 674Hz. Still, most researchers continue to consider 700Hz a benchmark for the raising of the THOUGHT-vowel, and I use this number in the analyses below.



Map 10.31. The relative height of /oh/ in *caught*, *cause*, *law*, etc.

Figure 45: Values for F1 of THOUGHT-values across North America, from the Atlas of North American English (2006:108)

Despite the fact that some raising occurs in other places along the eastern corridor, there is evidence that raised-THOUGHT has become strongly indexical of NYCE specifically (Becker 2010, Johnstone 2016). Its use as a NYCE feature has permeated the media and popular culture. A well-known example is a series of sketches on "Saturday Night Live" from the early 1990s called *Coffee Talk*. In this sketch, Mike Myers, who is Canadian himself, imitates his then-mother-in-law Linda Richman, a Jewish woman from Long Island. His impression of New York City speech heavily features raised THOUGHT, notably in both words of the title of the sketch.

6.1.3 A change in progress, but in which direction?

Based on his apparent-time findings, Labov (1966) argued that THOUGHT-raising was undergoing a change from below in NYCE towards higher realizations of the vowel. Evidence for his proposed change in progress included analysis of the variable's distribution across social classes. Labov found a curvilinear distribution, seen in Figure 46 below, in which the middle groups (upper-working-class and lower-middle-class) had the highest rates of THOUGHT-raising, indicated by a lower raising score on the y-axis (1966:164).

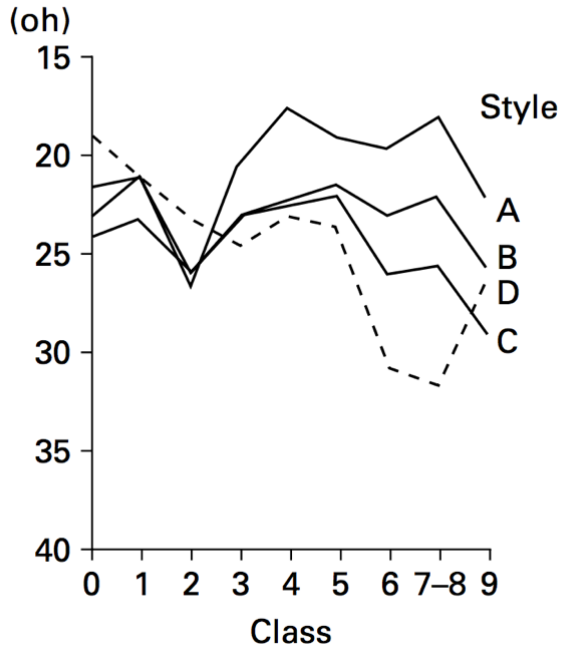


Figure 46: This graph from Labov 1966 shows the rate of THOUGHT-raising (indicated by the calculated raising scores on the y-axis) across different social classes, numbered from 1 (lowest class) to 9 (highest class). Each line represents a different style of language use, ranging from the least monitored (A) to the most monitored speech (D).

A curvilinear distribution for a variable across styles and social classes, like the one above, is evidence often used to support claims of change in progress (Labov 1972, Bailey 2002). Additional evidence to motivate Labov's claim is the cross-over pattern seen in Figure 46. Labov found hypercorrection among the lower middle-class speakers in more formal styles (indicated by the dip in the dotted line on the graph, occurring for social classes 6 and 7 in the most formal style). This provides further evidence for change in progress and indicates a linguistic insecurity for these speakers with regard to this particular linguistic feature (165).

The more recent ANAE continues to identify the raising of THOUGHT as a characteristic of NYCE (Labov et al. 2006) but Becker 2010's work in the same area of Manhattan as SSENYC (the Lower East Side), provides apparent-time evidence showing that the change towards raised-THOUGHT proposed by Labov in earlier literature is not borne out, and is in fact reversing among white speakers. While her oldest white speakers born before 1946 (both Jewish and non-Jewish) continued to raise their THOUGHT vowels, Becker found that few young white speakers were doing so, only one out of eight of the speakers born after 1970. This lowering in apparent time of THOUGHT for young, white speakers has also been found in other regions, notably Philadelphia (Labov et al. 2013).

6.1.4 Ethnicity and THOUGHT-raising

While the current study cannot contribute much to the scholarship on ethnicity and THOUGHT-raising, it is worth briefly reviewing the previous research in this area because it is an important part of understanding the social associations of this feature in the NYC area. Additionally, Jewish as an ethnicity has been at the center of many of these analyses (e.g. Labov

1966), and that is relevant to the population of the sample in this study, most of whom are Jewish.

Labov's 1966 work looks primarily at three ethnic groups in the Lower East Side that he deemed to have had large enough numbers of native English-speaking adults to study: Jewish, Italian, and African American. Labov found that the Jewish speakers in his sample showed a tendency towards a higher THOUGHT-vowel than did the Italian speakers (1966:186). His explanation for the higher THOUGHT pronunciation by Jewish speakers was a substratum effect of Yiddish. Since many L1 speakers of Yiddish did not distinguish the two low back rounded vowels from each other in English, Labov proposed that the children of these speakers were hypercorrecting as a way to distance themselves from their parents' pronunciation, and thus were producing overly raised THOUGHT-vowels (1966:193). Labov does not provide any evidence in the form of meta-linguistic commentary from young Jewish speakers indicating awareness of or linguistic insecurity about their parents' merged vowels, but it is worth noting that a 1979 study by Laferriere in Boston also found Jewish speakers to be raising THOUGHT higher than their Italian counterparts.

Labov 1966 generally excludes African Americans from his analyses of raised-THOUGHT because he found that they were not showing regular patterns of variation in regard to social class or speaking style (231). There is some usage of raised-THOUGHT by African Americans in Labov's speaker sample, but he concludes that "the gradual evolution of New York City speech towards higher (oh) forms has not been followed by AA speakers" (233). This conclusion, however, is at odds with the findings of more recent studies that have found African American speakers of all ages producing raised THOUGHT (Coggshall and Becker 2010, Becker 2010).

Members of other ethnic groups have also been shown to raise THOUGHT (see Wong 2007, 2015 for American born Chinese, and Slomanson and Newman 2004 for Latino adolescents).

6.1.5 Summary

Becker 2010 found that overall, THOUGHT has been lowering in apparent time (for Jews and other white speakers), directly contrary to Labov 1966's prediction of a change in process towards higher pronunciations of THOUGHT. Newman 2014, however, provides data from speakers in the outer boroughs who appear to be more conservative than the Lower East Siders who are the focus of Becker's study. This may be evidence that the reversal of the change in progress found by Becker is either not occurring outside of Manhattan or is progressing more slowly in the outer boroughs. It then stands to reason that Nassau County speakers of NYC English may also show a different pattern from Becker's Manhattan speakers. In the rest of this chapter, I will present evidence that THOUGHT-raising is indeed used by young speakers in Nassau County.

In section 6.2 I describe the methodology I used for measuring and analyzing the vowels. In 6.3 I analyze both the THOUGHT and LOT vowels of the speakers, and how the two vowel classes relate to each other in the vowel space of each speaker. Having established that all speakers in this study have distinct THOUGHT and LOT vowel categories (i.e. do not have the low back merger), I then turn in section 6.4 to the more salient aspect of this vowel for NYCE, an investigation of the amount of THOUGHT-raising that the speakers display.

6.2 Acoustic Measurement and Data Inclusion

I extracted all of the tokens of THOUGHT from each speaker's sociolinguistic interview and measured the formants at 33% into the vowel using FAVE. I excluded tokens where the vowel did not carry primary stress in the word, where the vowel was less than 50 milliseconds in duration, where it was preceded by an /l/ or a glide, or where it was followed by an /r/. Tokens with following /r/ were excluded because of the possibility for those /r/s to be non-rhotic, another feature of the NYC dialect that is not investigated in this study, but which is variably present amongst some speakers in the population. Tokens with preceding /l/ and glides are excluded because they pose an uncertainty as to where to segment the consonant from the following vowel, and this is particularly a concern with an automatic aligner. Since my measurements are taken at 33% through the vowel, however, I have not excluded tokens with *following* /l/, as the segmentation at the end of the vowel is not at issue with a measurement early in the vowel nucleus. All of my statistical analyses (MANOVA and regression) also account for the effects of preceding and following phonological environment. I included instances of the word *because* when I judged them to be fully pronounced and not reduced. This resulted in a total of 2340 tokens of THOUGHT and 4396 tokens of LOT, after outliers were removed.

6.2.1 Discrimination of THOUGHT tokens

As a native speaker of the dialect, I initially used my own intuitions to decide whether each word in the corpus belonged to the THOUGHT or LOT word class. To ensure the validity of these categorizations further, I referenced the extensive list compiled by Becker in Appendix 4 of

her dissertation (2010a:248). Becker's list is comprised of 307 words and judgments for the discrimination of each word from up to seven sources. Three of these sources were native speakers that she consulted, and the other four are reference volumes on sociophonetics (Kenyon 1924, Labov 1966, ANAE 2006, and Thomas 2001). Words on this list are highlighted in gray if they are "undecided," meaning that there is not agreement among the sources as to which word class the word belongs to. For example, it is not agreed on whether the word "on" (and derivatives like "upon") is in the LOT or THOUGHT class of words, so I excluded all tokens of these words from my data.

There were three words in my corpus that were excluded from Becker's analysis due to disagreement by her sources, but which I deemed to be defensibly THOUGHT words. These words are SONG, STRONG, and SOFT (and their derivatives). Table 10 shows the data about these tokens. The SONG class contains fourteen tokens of the words "song" and "songs," spoken by five speakers. The mean F1 for all fourteen tokens is 737Hz, and the mean for each speaker is below 800Hz. The STRONG class represents 23 tokens of the words "strong," "stronger," "strongest," and "Stronghold[16]." The mean of all 23 tokens is 736Hz. These tokens are produced by thirteen different speakers, with the highest mean F1 being 821Hz and the lowest being 577Hz. The SOFT class of words contains sixteen tokens of the words "softball," "softened," "softest," and "software." The mean F1 for all sixteen tokens is 727Hz. Of the eight speakers who produce these tokens, the mean F1s range from 556Hz to 844Hz. Table 10 summarizes the means for

[16]Stronghold was the name of the FIRST Robotics game that season.

each of these three words and compares them against the means in the whole corpus of THOUGHT and LOT tokens.

	SONG	STRONG	SOFT	Full Corpus THOUGHT	Full Corpus LOT
Mean F1	737Hz	736Hz	727Hz	694Hz	824Hz
Total Number of Tokens (N)	14	23	16	2340	4396

Table 10: Summary of the range of means for SONG, STRONG, and SOFT compared against the means for THOUGHT and LOT in the whole corpus.

While the Ns for each of these three word classes are small, the F1 values show that in this corpus they are considered part of the THOUGHT class by the speakers, and thus I include them in my analysis.

6.3 Distinction Analysis

In this section I present the results of MANOVA (multivariate analysis of variance) on each speaker's distribution of THOUGHT and LOT vowels. MANOVA allows analysis of more than one dependent variable at the same time and can account for internal factors influencing the production of merger, such as phonological environment. The dependent variables in these analyses are F1 and F2 and account for both preceding and following phonological environment. The relevant output of MANOVA is the Pillai-Bartlett statistic, which I refer to henceforth as the Pillai score. Pillai scores were introduced into the study of vowel mergers by Hay, Warren & Drager (2006), and have been shown by several sociophoneticians to be a better measure of

vowel merger than previously used measures like Euclidian distance (Hall-Lew 2010, Nycz and Hall-Lew 2013). The Pillai score is also the default statistic used in MANOVA in the R statistics analysis environment (Baayen 2008). The Pillai score ranges from 0 to 1, and a higher score indicates a greater distance between the two vowel pronunciations, including both the F1 and F2 dimensions.

6.3.1 MANOVA Results

The mentors are listed in Table 11 below in order from highest to lowest Pillai score. Also listed are the mean F1 and F2 for each speaker in each word class.

	Mean Formants for /ɔ/		**Mean Formants for /ɑ/**			
Speaker	**F1**	**F2**	**F1**	**F2**	**Pillai Score**	**Significance of MANOVA**
Mr. Spector	592	1010	836	1454	0.839	p<.001
Mr. Geissler	596	1045	829	1417	0.804	p<.001
Mr. Price	623	972	817	1401	0.747	p<.001
Lenny	594	913	793	1343	0.662	p<.001
Tony	679	1008	861	1278	0.660	p<.001
Isaac	612	1017	738	1317	0.617	p<.001
Nate	703	1050	836	1337	0.588	p<.001

Table 11: Mentors listed from highest Pillai score to lowest.

The MANOVAs reached a significance of p<.001 for each of the mentors, indicating that THOUGHT and LOT are separate phonemes in their vowel inventories. Indeed, as I will discuss, a

level of significance of p<.001 also characterized the MANOVAs of every one of the students. As an examination of Table 11 shows, despite all the mentors being unmerged, there is some variation in the degree of separateness of the two categories, especially in the younger speakers. Mr. Geissler (aged 60) has the highest Pillai score of the Long Island mentors[17], indicating the largest distinction between the two vowel classes, and Nate (30) has the lowest. I present Mr. Geissler's and Nate's vowel plots for all tokens of both THOUGHT and LOT in Figures 47 and 48 respectively.

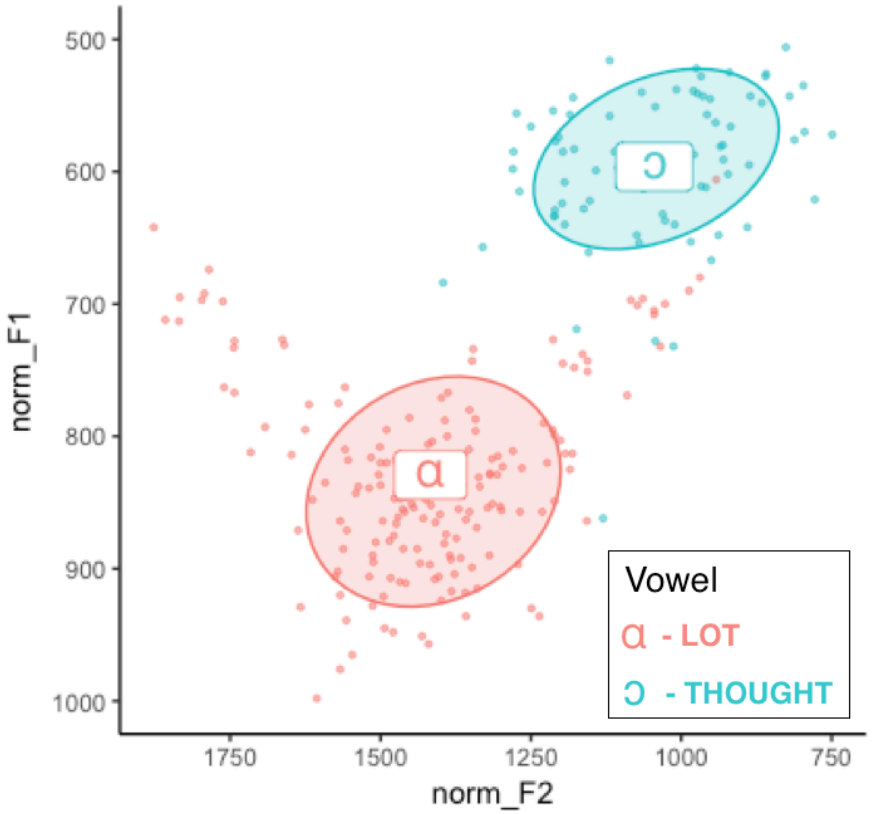


Figure 47: Plot of Mr. Geissler's THOUGHT and LOT tokens, illustrating a Pillai score of .804

[17]Recall that Mr. Spector and Lenny grew up in Brooklyn. In the statistical analyses in this chapter, I use only the five "Long Island mentors" as the group to contrast against the students, in order to find any change over time in the speech of Nassau County speakers specifically.

In Figure 47, the two low back vowels are very distinct. The means are more than 200Hz apart from each other, and the ellipses (representing one standard deviation) are small, showing tightly clustered data. Mr. Geissler's mean FI for THOUGHT (/ɔ/) is 596Hz and his mean F1 for LOT (/ɑ/) is 829Hz. The graph also illustrates that almost all of Mr. Geissler's tokens of THOUGHT are raised and most have F1 values of less than even 600Hz. Raising of the THOUGHT vowel will be discussed further in the next section.

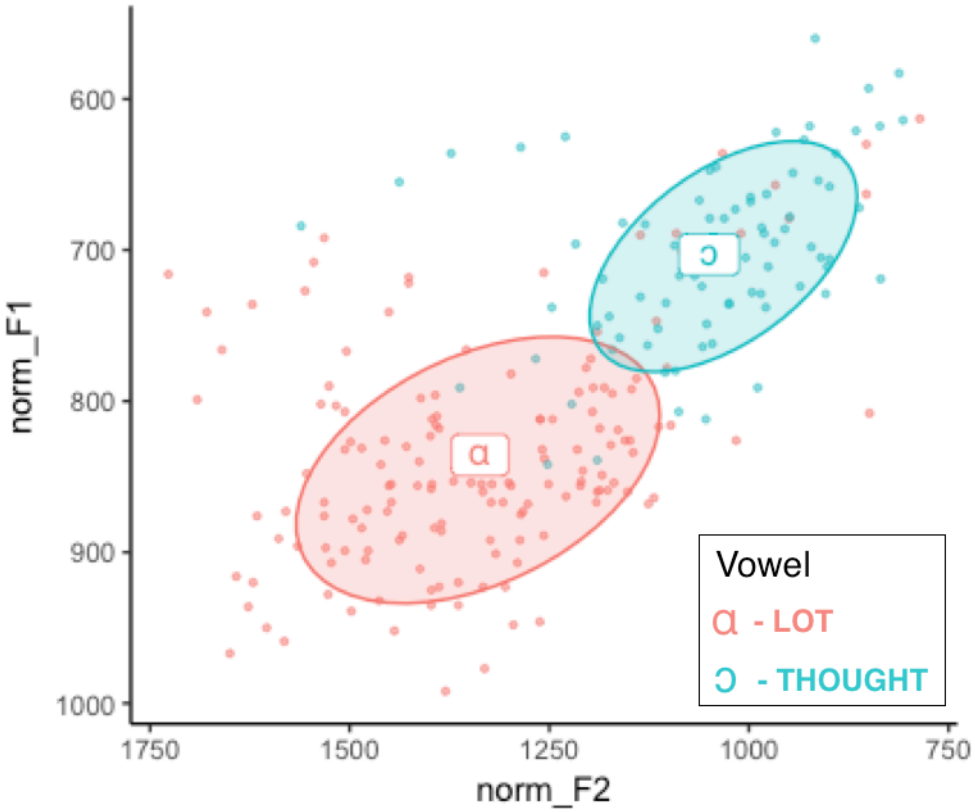


Figure 48: Plot of Nate's THOUGHT and LOT tokens, illustrating a Pillai score of .588

In contrast to Mr. Geissler's vowels, Figure 48 shows that Nate's two vowel classes are less distinct. They only overlap to a very small degree, but still look quite different from the large gap between the two categories seen in Mr. Geissler's graph. Nate's mean F1 for THOUGHT is 703Hz and his mean F1 for LOT is 836Hz. This is a difference of 133Hz, significantly less than Mr. Geissler's difference of 233Hz.

The following two tables show the same measurements for each of the 24 students, with Table 12 listing female students, and Table 13 listing male students.

	Mean Formants for /ɔ/		**Mean Formants for /ɑ/**			
Speaker	**F1**	**F1**	**F1**	**F2**	**Pillai Score**	**Significance of MANOVA**
Tahani	714	1011	869	1285	0.632	p<.001
Amy	775	1115	852	1370	0.588	p<.001
Chrissy	704	1099	841	1311	0.567	p<.001
Leah	709	1104	826	1332	0.507	p<.001
Diviya	696	1139	802	1313	0.413	p<.001
Becky	731	1147	836	1315	0.396	p<.001
Carina	686	1089	785	1276	0.388	p<.001
Chloe	753	1118	817	1289	0.264	p<.001
Hannah	735	1105	786	1222	0.229	p<.001

Table 12: Female Students listed from highest Pillai score to lowest.

	Mean Formants for /ɔ/		Mean Formants for /ɑ/			
Speaker	**F1**	**F1**	**F1**	**F2**	**Pillai Score**	**Significance of MANOVA**
Daniel	586	1033	847	1410	0.811	p<.001
Logan	685	1089	856	1364	0.688	p<.001
Caleb	681	1127	826	1377	0.682	p<.001
Matan	705	1031	830	1341	0.666	p<.001
Ethan	708	1066	838	1432	0.633	p<.001
Adam	667	1129	846	1352	0.619	p<.001
Joe	694	1099	835	1412	0.592	p<.001
Garrett	755	1121	871	1382	0.583	p<.001
Paul	682	1093	815	1338	0.551	p<.001
Tim	728	1065	815	1321	0.517	p<.001
Luke	726	1108	831	1324	0.498	p<.001
Jake	680	1179	807	1394	0.483	p<.001
Ranjit	722	1111	787	1321	0.412	p<.001
Alex	765	1151	825	1328	0.386	p<.001
Noah	738	1155	800	1357	0.330	p<.001

Table 13: Male Students listed from highest Pillai score to lowest.

The range of Pillai scores is much wider for the students than it was for the mentors. The highest Student Pillai score is Daniel with .811, which is exceeded by only one of the mentors (Mr. Spector). Daniel is the only student with a score this high, though, and the majority of students have scores in the range of .4 - .7. The lowest Pillai score is that of the student Hannah, who has a score of .229. Even with this low score, the MANOVA reaches significance at p<.001,

meaning that Hannah does have two separate vowel classes, but they overlap to a higher degree than those of her peers. To illustrate this, Figures 49 and 50 show the distribution of THOUGHT and LOT vowels of the two students on either end of the spectrum: Daniel and Hannah. Both students are white and Jewish; Daniel is a senior and Hannah is a junior.

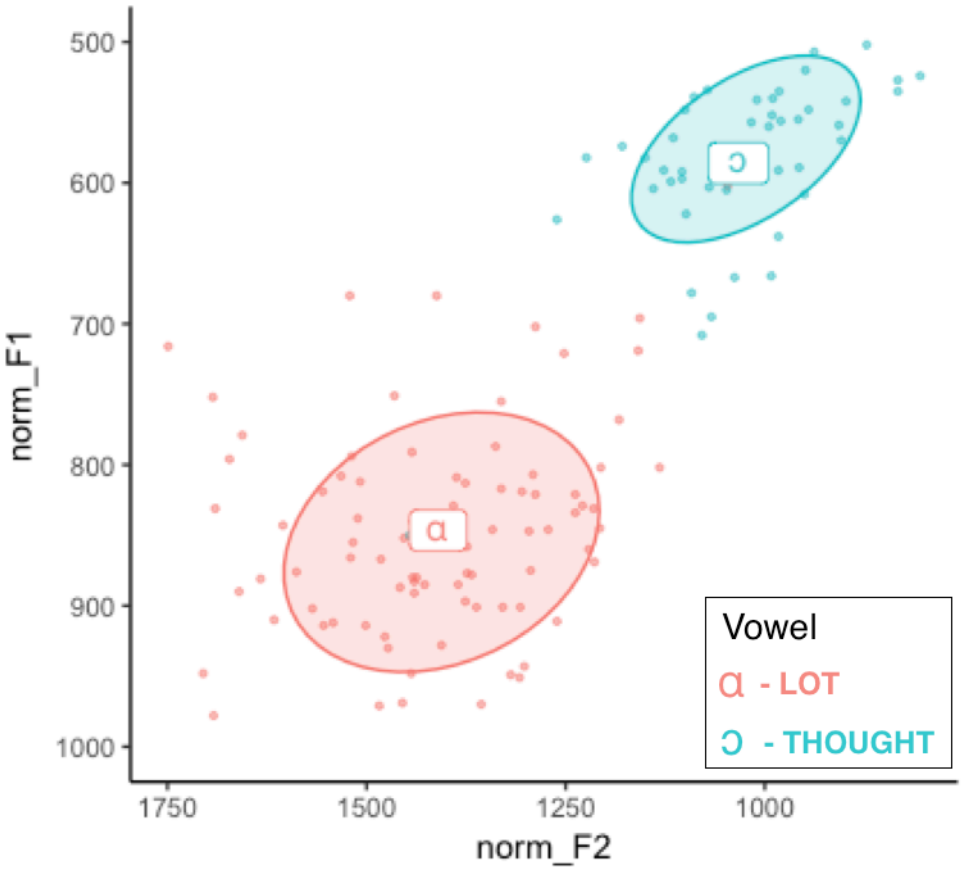


Figure 49: Plot of Daniel's THOUGHT and LOT tokens, illustrating a Pillai score of .811

In Figure 49, the two low back vowels are clearly very distinct, similar to those of Mr. Geissler. The means are far away from each other, and the ellipses (representing one standard deviation) are relatively small, showing tightly clustered data. His mean FI for THOUGHT (/ɔ/) is

586Hz, and his mean F1 for LOT (/ɑ/) is 847Hz, a difference of 266Hz. It is also visible on this graph just how raised Daniel's THOUGHT vowels are. In fact, Daniel has the highest THOUGHT tokens of all the students by a wide margin. His mean value for F1 of THOUGHT is even lower than that of any of the mentors. Almost all of Daniel's THOUGHT tokens are below 600 Hz, while the majority of his LOT tokens are at 800Hz or above, leaving a wide gap between the two categories.

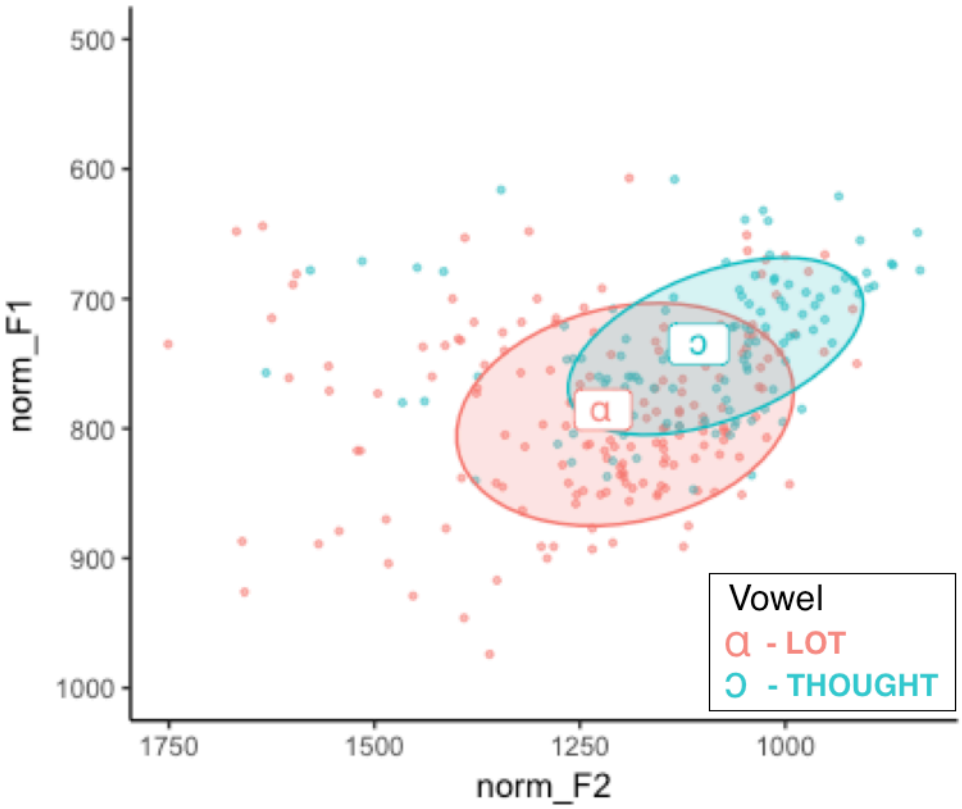


Figure 50: Plot of Hannah's THOUGHT and LOT tokens, illustrating a Pillai score of .229

In contrast to Daniel's plot above, Hannah's vowel distribution seen in Figure 50 shows the most overlap between the two categories of any of the students. The ellipses for THOUGHT

and LOT overlap a fair amount, showing a less stark distinction between the two categories. The ellipses themselves are also bigger, showing larger variation in her production. Hannah's mean F1 is 735Hz for THOUGHT and 786Hz for LOT, leaving only a difference of 51Hz. Despite this degree of overlap, the means are still significantly different, and Hannah does not have a merger or near merger of these two vowels. Also seen on the plot is that some of Hannah's THOUGHT tokens are indeed below 700Hz, indicating that she does actually raise the THOUGHT vowel sometimes. The amount of raising each speaker uses is discussed further in section 6.4.

6.3.1 Regression Analysis

Because the calculation of the Pillai score as part of a MANOVA has already taken into account the phonological environment, each speaker's Pillai score can be used as the dependent variable in a model testing for the effects of social factors on the degree of vowel class distinction. I performed a fixed effects regression model in R[18] with the dependent variable of Pillai score and the independent social variables of gender, age, and ethnicity. This analysis was performed only on the student data, as the mentors do not vary in gender or ethnicity, and so cannot be included in an analysis of these variables. The full results of the resulting model are shown in Table 14.

[18]Model=lm(PillaiScore ~ Gender + Age + Ethnicity, data = StudentVowels)

	Estimate	Standard Error	t-value	p-value	Significance
(Intercept)	-0.43826	0.36850	-1.189	0.2483	
Gender (Male)	0.15454	0.05944	2.600	0.0171	p<.05
Age	0.05832	0.02245	2.593	0.0174	p<.05
Ethnicity (White)	-0.08325	0.07144	-1.165	0.2576	Not Sig

Table 14: Full results of linear fixed effects model of social factors on Pillai scores of students

The model shows that there are significant effects of both gender and age, although the effect sizes (estimates) are extremely small. Both of these effects reach significance at p<.05. The analysis shows that the male students have slightly higher Pillai scores, corresponding to slightly larger distinctions between their THOUGHT and LOT vowel classes. In addition, there is a small effect of age, with older students having higher Pillai scores than younger students.

There is no significant effect of ethnicity in the amount of overlap of vowel classes, however there is not very much data in the sample to base it on. Looking at the ranking of student Pillai scores in the above tables, it can be seen that the Asian speakers are not clustered at the bottom but are distributed throughout the white students. This finding is counter to a claim in Newman 2014 that suggested South Asian speakers in Queens have a complete low back merger and show no significant difference between the THOUGHT and LOT vowel classes. The three Indian students in this study (Tahani, Diviya, and Ranjit) are distributed throughout the range of possible scores, with Tahani actually having the highest Pillai score of all the female students.

6.4 Raising Analysis

In this section I turn to an analysis of the amount of THOUGHT-raising in the speaker sample. Unless otherwise noted, all plots in this section have a y-axis that ranges from 500Hz to 900Hz so that it is easier to compare the height of tokens between plots visually. I at first attempted to divide the speakers into "Raisers" and "Non-Raisers" based on the 700Hz threshold, but this proved to be an unenlightening endeavor, as every speaker in the sample has some tokens that are raised, and there is also wide variation in how high those tokens are raised. So instead, I illustrate each speaker's THOUGHT tokens with boxplots that visually represent the distribution of their THOUGHT production, with a blue line drawn at 700Hz on each plot for reference. I also list each speaker in a table with their Mean F1 for THOUGHT, and also the percentage of their THOUGHT tokens that are below 700Hz, as well as a column that shows how many are extra-raised, with an F1 lower than 600Hz. In section 6.4.1 I present the data for the mentors, and in 6.4.2 the data for the students. Finally, I investigate whether any social factors impact the relative amounts of raising among speakers.

6.4.1 Mentors

The mean F1 for all seven mentors as a group is 629Hz, a value well below the threshold of 700Hz as a benchmark for THOUGHT-raising. Removing the two Brooklyn-born mentors raises the average F1 a little bit to 643Hz. Figure 51 shows the boxplots for each mentor's production of THOUGHT and Table 15 lists information about percentage of tokens that are raised, along with each speaker's demographic information.

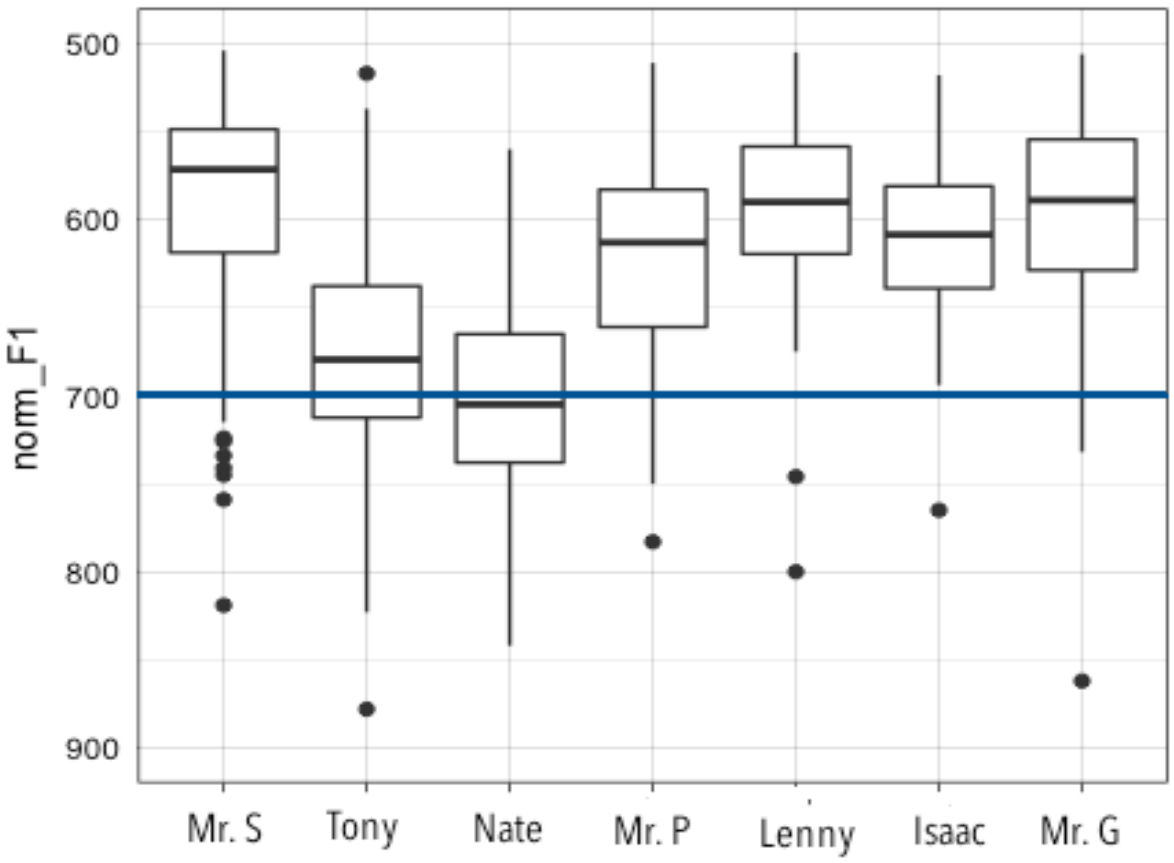


Figure 51: Boxplots of the Mentors' F1s for THOUGHT

Subject	Under 700 Hz	Under 600 Hz	Avg F1	N		Age	Gender	Ethnicity	Religion
Isaac	99%	35%	612	78		59	Male	White	Jewish
Lenny	97%	56%	594	59		86	Male	White	Jewish
Mr. Geissler	95%	60%	596	78		65	Male	White	Catholic
Mr. Spector	92%	66%	591	104		66	Male	White	Jewish
Mr. Price	88%	42%	623	85		30	Male	White	Catholic
Tony	65%	9%	679	88		24	Male	White	Catholic
Nate	48%	4%	703	81		31	Male	White	Jewish

Table 15: Mentors listed from highest percentage of raised THOUGHT to lowest.

The mentor Isaac has virtually all of his tokens below 700Hz, but only 35% of his tokens are below 600Hz. While Lenny, Mr. Geissler, and Mr. Spector have slightly lower overall percentages of raised tokens, their tokens that *are* raised are much more often under 600Hz than for Isaac. Nate's status among the mentors mirrors what was seen above in section 6.3.1, where Mr. Geissler had the highest Pillai score and Nate the lowest.

Figure 52 shows the five Long Island mentors plotted by mean F1 and age. The best fit trendline shows a relatively strong correlation between age and mean F1, with the older mentors raising more than the younger mentors.

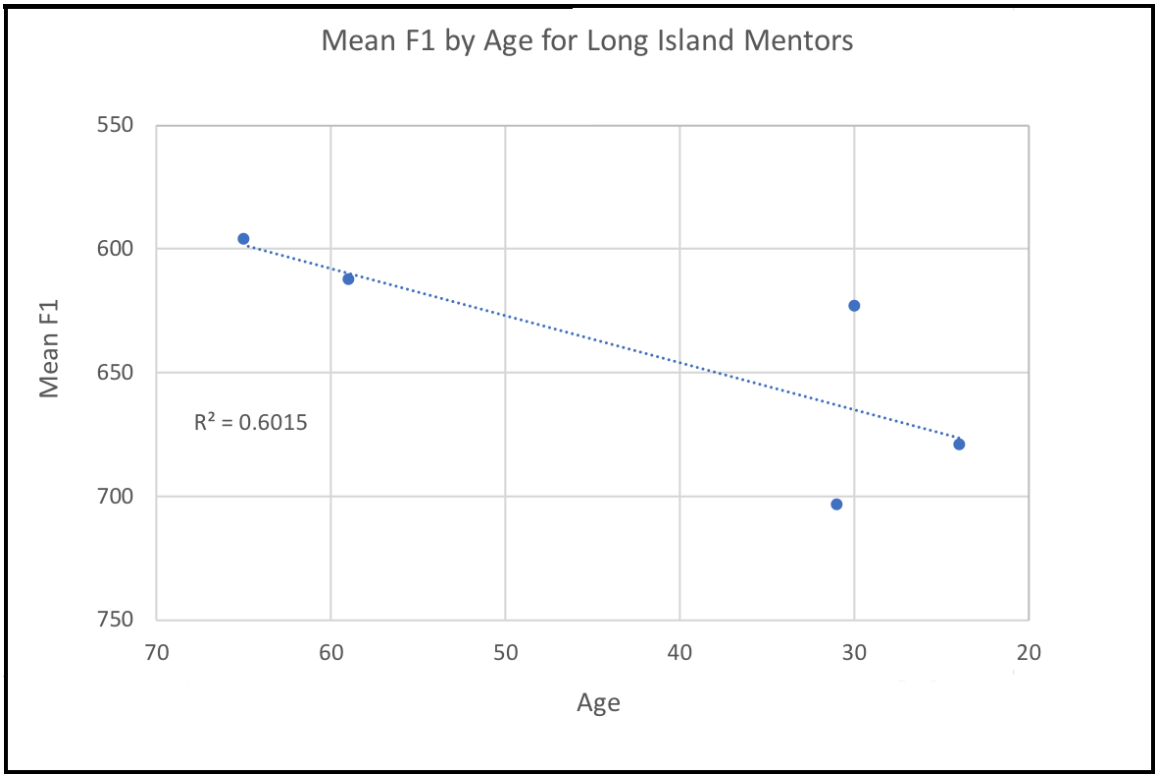


Figure 52: Plot of mean F1 for THOUGHT in Hz by age for the five Long Island Mentors, with a best fit line.

6.4.2 Students

I have divided the students into two sets of plots, each containing twelve students, in order to make the information possible to read and compare. First, I present the top half of students in regard to percentage of tokens with an F1 under 700Hz, followed by the bottom half. This delineation is purely for ease of reading the graphs and is not indicative of any actual divide that takes place at this point in the distribution; the last member of the top half, Chrissy, has 47% raised tokens, while the first student on the second list has 46% raising. Figure 53 shows boxplots for the first twelve students, and Table 16 lists each of these student's mean F1, percent of tokens under 700Hz, under 600Hz, number of THOUGHT tokens (N) as well as their social characteristics.

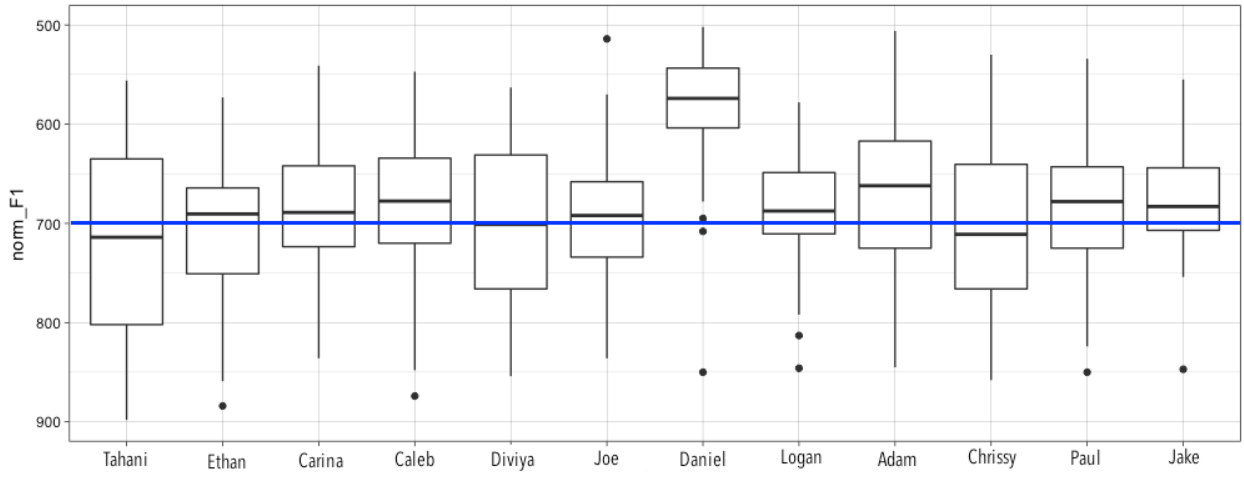


Figure 53: Boxplots of THOUGHT F1s for the first half of students

Subject	Under 700 Hz	Under 600 Hz	Mean F1	N		Age	Gender	Ethnicity	Religion
Daniel	96%	72%	584	46		Senior	Male	White	Jewish
Jake	71%	10%	680	21		Sophomore	Male	White	Jewish
Adam	67%	15%	661	39		Junior	Male	White	Jewish
Logan	61%	7%	685	56		Junior	Male	White	Jewish
Caleb	61%	8%	679	142		Freshman	Male	White	Jewish
Carina	56%	9%	682	64		Junior	Female	Chinese	
Ethan	56%	1%	708	70		Sophomore	Male	White	Jewish
Paul	55%	12%	680	49		Senior	Male	White	Jewish
Joe	55%	8%	694	77		Junior	Male	White	Jewish
Tahani	49%	11%	714	53		Sophomore	Female	Indian	Sikh
Diviya	49%	11%	696	37		Freshman	Female	Indian	Sikh
Chrissy	47%	10%	704	87		Senior	Female	White	Catholic

Table 16: Students ranked from highest to lowest percent of raised THOUGHT tokens

Daniel, who had the highest Pillai score, also does the most raising of any of the students by far. 72% of his tokens were under 600Hz. The students are listed in order of percent of raised tokens, but you can see that their mean F1s do not follow in the same order. Ethan, for example, has a mean F1 of 708Hz, which would have put him over the threshold for being "a Raiser" had I used such a cut off, but he actually raised 56% of his THOUGHT tokens, and this is a better measure of how "New York" he actually sounds. Tahani also has a mean F1 over 700Hz, but she raises almost half of the time, illustrating the problem with only looking at means in comparing the amount of THOUGHT-raising that speakers produce.

Figure 54 shows boxplots for the second half of the students, and Table 17 lists all of the same information for each of them as for the above students.

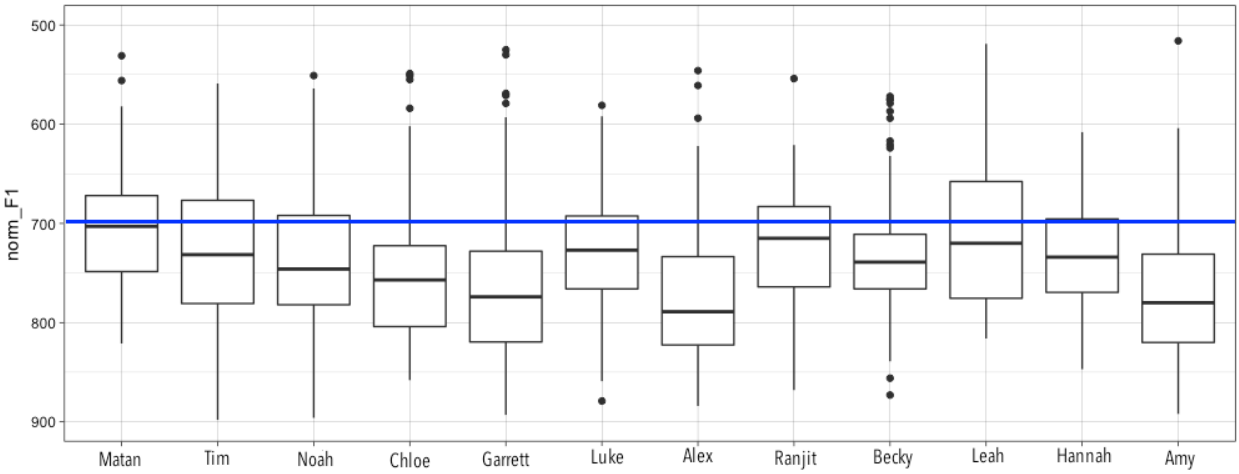


Figure 54: Boxplots of THOUGHT F1s for the second half of students.

Subject	Under 700 Hz	Under 600 Hz	Mean F1	N		Age	Gender	Ethnicity	Religion
Matan	46%	7%	705	91		Senior	Male	White	Jewish
Leah	44%	13%	709	48		Junior	Female	White	Jewish
Ranjit	41%	1%	722	69		Freshman	Male	Indian	Hindu
Luke	32%	3%	726	75		Junior	Male	White	Jewish
Tim	32%	2%	728	94		Freshman	Male	White	Jewish
Hannah	29%	0%	735	115		Junior	Female	White	Jewish
Noah	26%	4%	738	109		Sophomore	Male	White	Jewish
Alex	22%	6%	765	51		Freshman	Male	White	Jewish
Garrett	21%	8%	755	75		Senior	Male	White	Jewish
Becky	21%	5%	731	121		Freshman	Female	White	Jewish
Chloe	15%	4%	753	103		Freshman	Female	White	Jewish
Amy	13%	1%	775	75		Senior	Female	Chinese	

Table 17: Students ranked from highest to lowest percent of raised THOUGHT tokens

Figure 54 contains far more students whose mean F1 exceeds 700Hz than in Figure 53, but as mentioned before, there is a trend but no definitive cut off when a speaker shows no raising. The students with the least amount of raising are Chloe with 15% and Amy with 13%. Even though they do not raise THOUGHT most of the time, it is still important to note that it is in their linguistic repertoire and sometimes deployed.

6.4.3 Regression Analysis

For just the student data, I performed a fixed effects regression model in R[19] with percentage of a speaker's THOUGHT tokens that have an F1 below 700Hz as the dependent variable and independent variables of age, gender, and ethnicity. The results of this model show no significant effects of any of the social factors. Age, gender, and ethnicity do not correlate with the frequency of THOUGHT-raising of the student speakers. The full results of the model are presented below in Table 18.

	Estimate	Standard Error	t-value	p-value	Significance
(Intercept)	-2.423	36.881	-0.066	0.948	
Age	4.009	3.502	1.145	0.266	n.s.
Gender (Male)	15.644	9.273	1.687	0.107	n.s.
Ethnicity (White)	-6.412	11.146	-0.575	0.571	n.s.

Table 18: Full results of linear fixed effects model, showing no effect of social factors on percent of THOUGHT-raising among students.

[19]Model=lm(PercentRaised ~ Gender + Age + Ethnicity, data = StudentVowels)

Below in Figure 55 the student tokens of THOUGHT are plotted by gender, and the ellipses almost completely overlap, visually illustrating the lack of significance of gender on THOUGHT-raising on this population.

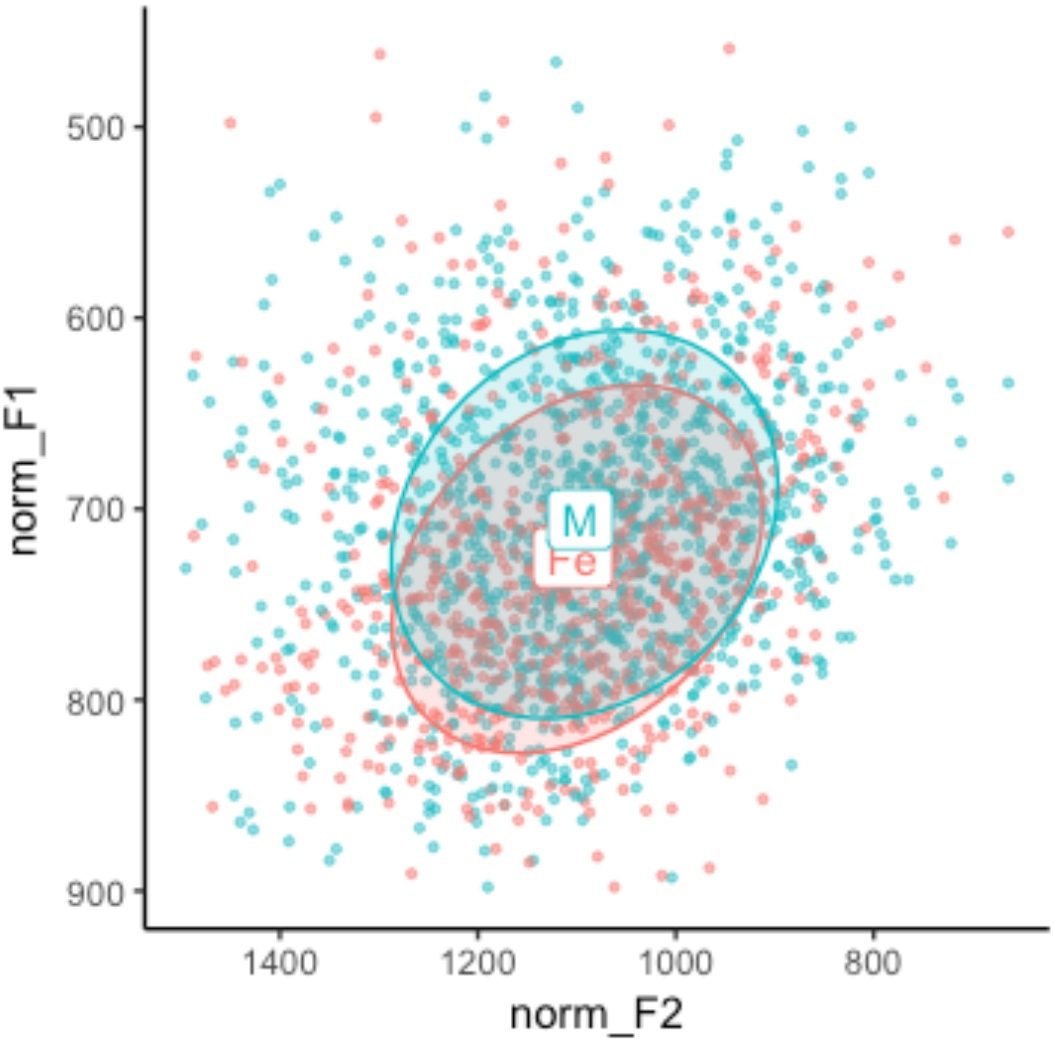


Figure 55: Student tokens of THOUGHT plotted by gender

6.4.4 Generation Gap

While there were no effects found in this sample of gender or ethnicity, age is somewhat a factor in the amount of THOUGHT-raising of the speakers. The Long Island mentors showed a significant effect of age in the trendline, while no such effect arose in the student model. However, the mentors span a much larger age range; the students are all within four years of age of one another, and so it is not surprising that no correlation emerges. The Long Island mentors as a group versus students as a group does yield a significant difference in THOUGHT-raising. This together with the correlation by age in the mentors, shows a raising of F1 (hence a lowering of the vowel) in apparent time. In Figure 56, all of the THOUGHT tokens in this analysis are plotted, grouped by (Long Island) mentors vs. students.

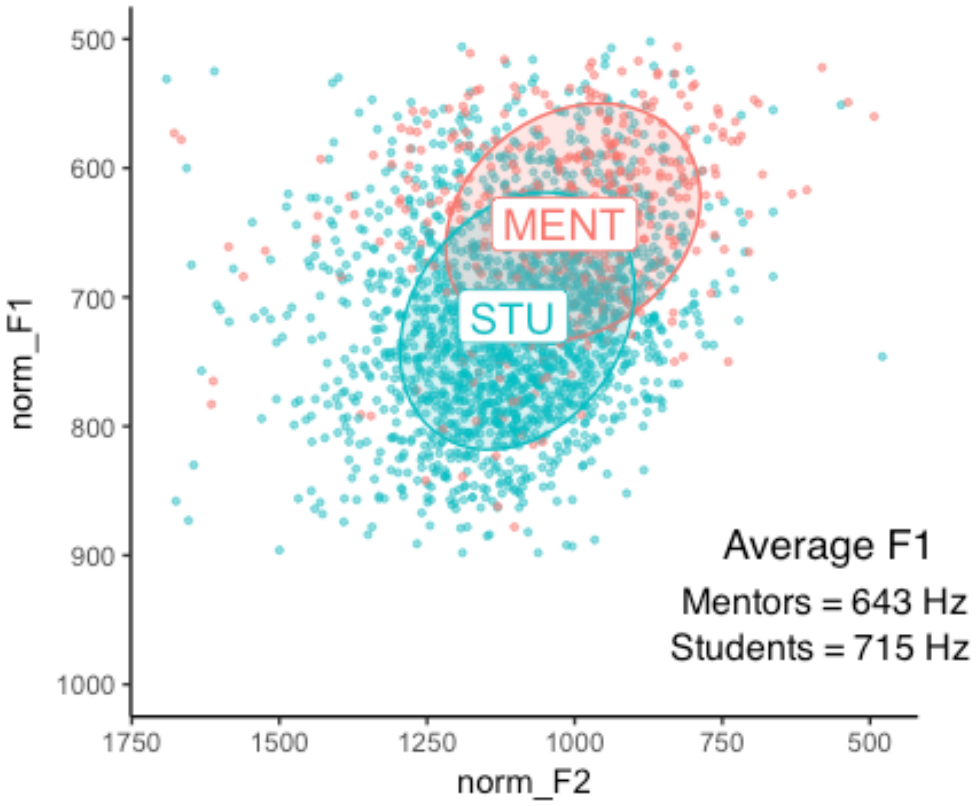


Figure 56: Vowel Plot of all THOUGHT tokens for Long Island Mentors and Students

The Long Island mentors' average F1 is 643Hz and the students' average F1 is 715Hz. While there is overlap of the ellipses, a MANOVA shows that there is a significant difference between the two groups at the level of $p<.001$. The MANOVA uses both F1 and F2 as dependent variables and so is accounting for both raising and backing of the vowel. The Pillai score calculated for the difference between the mentor THOUGHT vowels and the student THOUGHT vowels is .148.

6.5 Discussion

The goal of this chapter was to analyze Nassau County speakers' production of the low back vowels. To provide a complete picture of the realization of low back vowels, Sections 6.2 and 6.3 presented the aggregated patterns of the low back distinction and of raised-THOUGHT of the sample of the 31 speakers. The distinction analysis showed that all speakers, both mentors and students, maintained a significant division between the two low back vowels, but they display a range of degree of overlap, measured in Pillai scores. A regression analysis of the Pillai scores found small effects of gender and age on the vowel systems of the students. Male students were slightly more likely to have a larger separation between the two vowel classes than female students, and older students had slightly larger separations than younger students. Mentors exhibited larger separations between the two vowels than students did. Although there was a pattern of the phonemic space between THOUGHT and LOT decreasing in apparent time (as also found in Becker & Wong 2012), all speakers still had robust distinctions between the low back vowels, indicated by relatively high Pillai scores.

The analysis of raising showed that all speakers in the sample produce raised THOUGHT to some degree. Using a benchmark of 700Hz, there were no speakers who did not have any tokens of THOUGHT considered to be raised. The speakers did vary in how much and how high they raise, and looking at the whole distribution rather than just mean F1 is important in gauging the amount of raising a speaker produces. Almost all of the mentors are very high raisers. The mentor Nate had the least amount of raising and he also had the lowest Pillai score of the mentors. He is the same mentor who differed from the rest in regard to short-*a.* Nate is the

second youngest mentor, and he shows the least NYCE features out of the mentors in terms of both of the vowels analyzed here.

There is an overall significant effect of age among the Long Island mentors, with older speakers having lower mean F1s and therefore showing a higher degree of THOUGHT-raising. This age pattern continues in a comparison between the mentors and students as generational groups; a MANOVA shows that the mentors raise and back THOUGHT significantly more than students do.

Among the students there is a wide range of degrees of raising. An analysis of how often the students raise their THOUGHT tokens, and how high they are raised, showed that while there is a wide range of variation, all of the students raise THOUGHT to some degree. This is surprising given recent reports of THOUGHT-lowering among young white New Yorkers in Manhattan (Becker 2010). The fact that these students, born in the years 1998 through 2002, raise THOUGHT as much as they do is noteworthy.

It is also noteworthy that there are no social patterns explaining which students raise and which ones do not. Ethnicity has been found to be a significant factor in recent studies of THOUGHT in New York City, but in this sample, it did not play a role in determining rates of raising. This result must be tempered, however, because of the small number of students in the current sample who are not in the majority group of white Jews. It is noteworthy still that, while the social factors of gender and ethnicity *did* correlate with the likelihood of a student having a Tragerian short-*a* system, no such patterns arose in the THOUGHT data.

The one social factor that was found to have an effect on THOUGHT-raising was age on a generational scale - between the mentors and the students, and between the older and younger

mentors. This is in line with the change in apparent time found by Becker on the Lower East Side, where younger speakers were progressively lowering THOUGHT over time. It adds to her evidence of the reversal of the change from below found by Labov (1966), which predicted that white, and Jewish speakers specifically, would continue to raise THOUGHT over time. But Becker's youngest speakers were significantly older than the teenagers in this study, and even ten to twenty years older than the young group of mentors. Only one of Becker's "young" speakers used raised-THOUGHT at all, while the frequency and extent is much higher in this sample of Nassau County teenagers. This may be evidence that the change in progress of THOUGHT-lowering in New York City English is progressing more slowly in areas outside of Manhattan.

Chapter 7: Conclusion

7.1 Summary of Results

7.1.1 Short-*a*

My investigation into the short-*a* split in Nassau County revealed three main results. The first is that the Tragerian short-*a* split certainly is still in use, even by some speakers as young as fourteen. The second broad finding is that there does seem to be a change in progress, à la Becker 2010, towards the loss of the Tragerian complex split in favor of the nasal split that is common across the United States. The third main finding concerns the two transitional states that were found in between the Tragerian split and the nasal split. The one I call Transitional marks the change of the BASH class from *tense* to *lax*, while the speakers retain the BAD class as *tense*. This seems to be a leading indicator of the process underway of simplifying the complex Tragerian split towards a nasal split. The other transitional stage is the one that I label NYC-Nasal. In this system, a basic nasal split surfaces but speakers retain the Open Syllable Constraint, a holdover of the full Tragerian system.

I am not the first to note the *laxing* of BASH by NYCE Speakers. Coggshall 2017 also found this among her population of speakers in Jersey City and identified it as possibly the first real indicator of a change in progress that is not tied to any inherent variation that had previously been described as part of the complex system. In addition, a recent NWAV paper by Haddican et al. (2018) reported that their data revealed a pattern among younger non-white speakers of ultra-

lax realizations of short-*a* in the pre-voiceless fricative context. In addition, Haddican et al. showed evidence that the shift in New York is happening piecemeal, in separate processes or stages. This runs counter to Labov 2016, which claims that a similar shift in Philadelphia (from a complex split to a nasal split) is happening as one wholesale change with no intermediate stages. My results agree with those of Haddican et al., in that I found both speakers *laxing* BASH while retaining *tense* BAD, as well as speakers who have an otherwise nasally conditioned split maintaining the Open Syllable Constraint.

7.1.2 THOUGHT

In regard to the THOUGHT vowel, one main finding here was a confirmation that all speakers in the sample, even the youngest, had a significant distinction between the low back vowels THOUGHT and LOT. However, there was a range in how separate the categories are for speakers, which I measured by Pillai scores.

More salient to the question of New York City English is the investigation of THOUGHT-raising. Though it has found to be in decline in Manhattan (Becker 2010) to the point that it might be expected to be totally lost from the repertoire of my youngest subjects, I found that the teenagers in Nassau County do still employ raising, albeit to differing degrees. Many of the students have a mean F1 for production of THOUGHT of less than 700Hz, and all of them produce tokens of raised THOUGHT to some degree.

The only social factor found to affect the production of THOUGHT in this sample is a broad measurement of age. There was a significant difference between mentors and students, and also among the mentors. My results are contrary to Labov's (1966) prediction that THOUGHT

would keep raising over time and are instead aligned with those of Becker 2010, who predicted a change in progress away from THOUGHT-raising. However, it seems from the rate of raising that is still used by the young speakers in my study, that the shift Becker found in Manhattan is not as far progressed in Nassau County.

7.1.3 Comparing the Two Sets of Results

In my analysis of the variation of short-*a* systems present in the population, I found evidence of social conditioning by gender and also by ethnicity, but with the THOUGHT vowel this was not the case. For short-*a,* only males displayed Tragerian or Transitional systems, while all of the female students had either a Nasal split or an NYC-Nasal split. Interestingly, though, between the two nasal systems, female students dominated in the NYC-Nasal category, meaning that they retained the Open Syllable Constraint more often than male speakers. While it is difficult to glean too much from the finding of an effect of ethnicity on short-*a,* due to the small number of minority students in the study, it remains that all five of the Asian students (Chinese and Indian) had nasal systems.

In regard to the THOUGHT-vowel, there was an effect of age supporting a change in progress towards lowered THOUGHT, but otherwise I did not find any social conditioning of THOUGHT-raising. One difference between the two variables that may be relevant is that THOUGHT-raising is a feature very much above the level of consciousness, noticed by people, and linked to NYC in an indexical way, while short-*a* is a feature that is under the radar and not very salient to speakers.

Despite this mismatch in social conditioning, there is a correlation between which speakers leading the changes in both paradigms. I performed two correlation tests that I will illustrate below. The first uses Pillai score to represent a speaker's distinction between the THOUGHT and LOT vowels, and perhaps their advancement in a move towards a low back merger, and in the second test I used the percentage of the speaker's THOUGHT tokens that are raised (have an F1 below 700Hz). In both cases, to represent short-*a*, I used a speaker's Tensing Score for words in the BAD class (in which short-*a* is followed by a voiced stop in a closed syllable). I chose this particular word class because it is the most accurate diagnostic of maintenance of a Tragerian split system. As shown in Chapter 5, the other characteristic word class, BASH, seems to be undergoing change, possibly in transition away from the classic complex system, but BAD remains resolutely part of a Tragerian split. Plots of these two correlations are presented below in Figures 57 and 58.

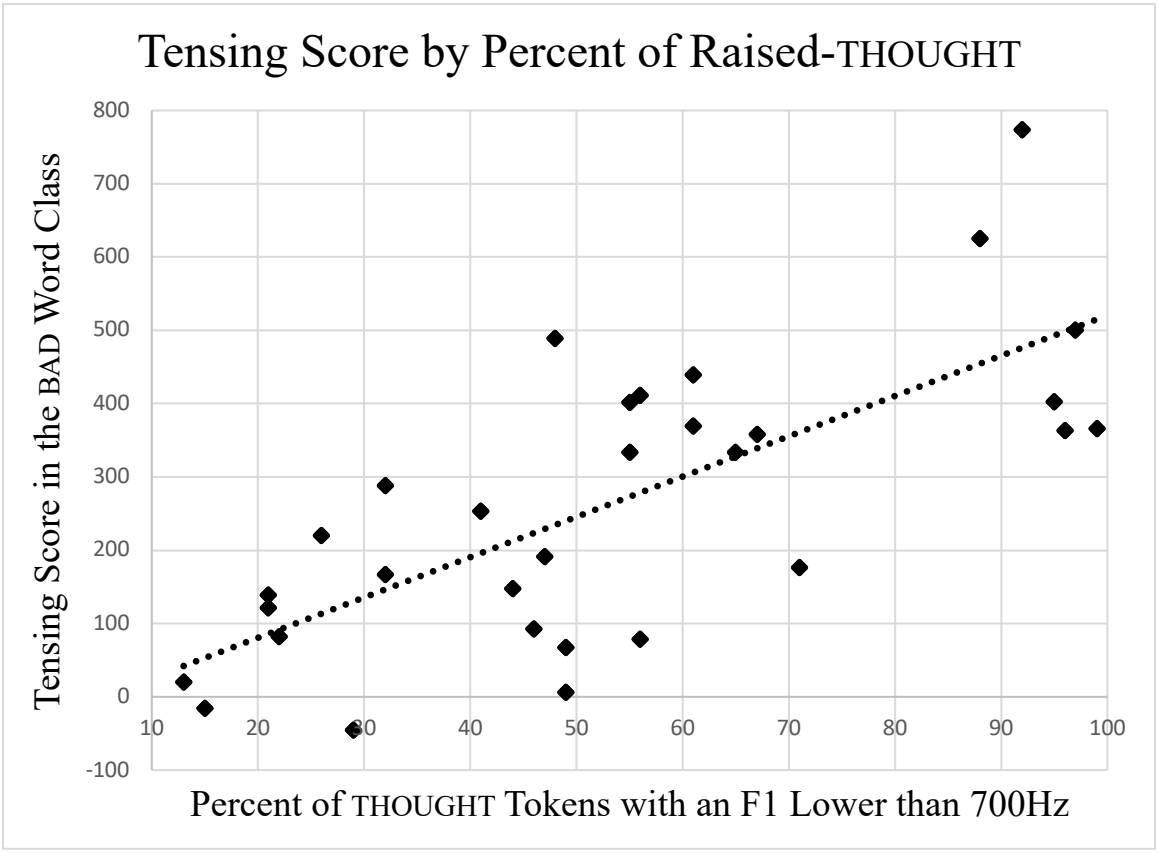


Figure 57: Chart showing the relationship between each speaker's percentage of raised-THOUGHT and their tensing score for words in the BAD word class.

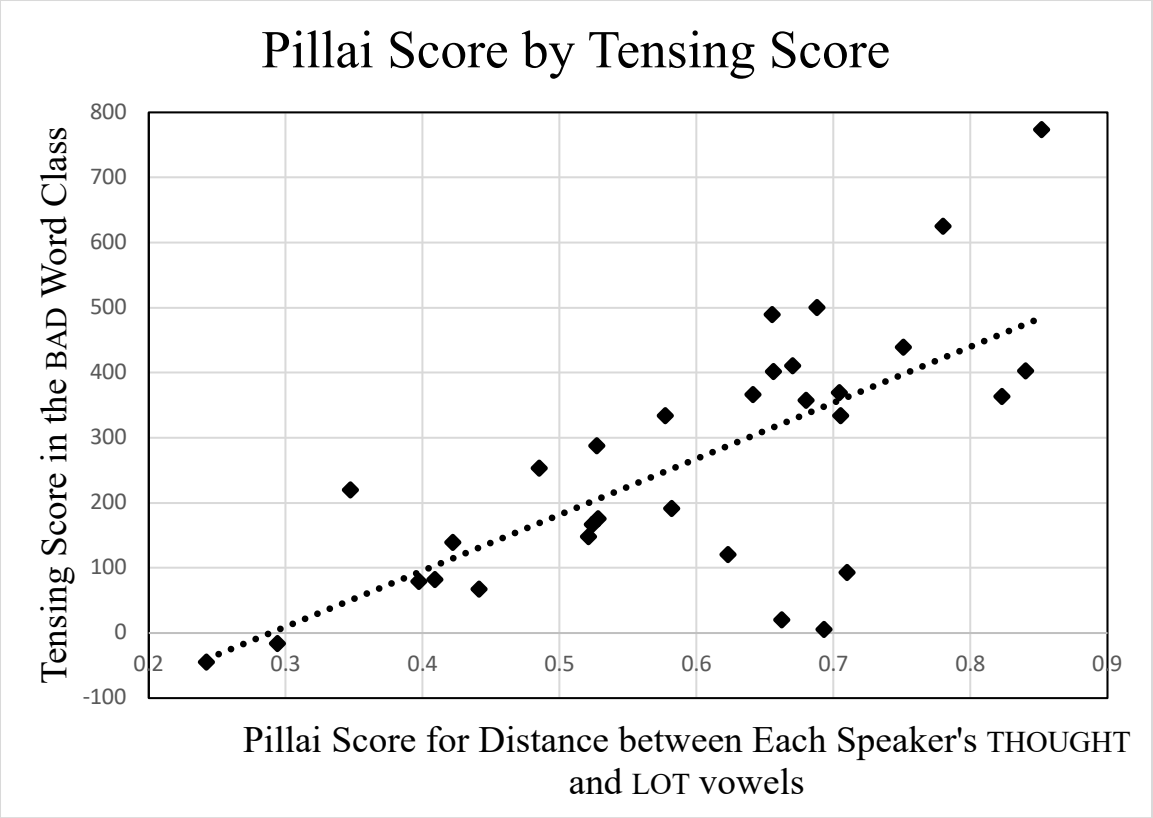


Figure 58: Chart Showing the Relationship Between Each Speaker's Pillai Score and their Tensing Score for the BAD Word Class.

Figure 57 shows each speaker's percentage of THOUGHT tokens that are raised (have an F1 below 700Hz) on the x-axis, and the speaker's Tensing Score for specifically the BAD word class (short-*a* with a following voiced stop in a closed syllable) on the y-axis. The best fit trendline represents a Pearson product-moment correlation coefficient of 0.6891 with p<.0001

Figure 58 shows each speaker's Pillai Score (representing the distance between their THOUGHT and LOT vowels) on the x-axis, and the speaker's Tensing Score for the BAD word class on the y-axis. The best fit trendline represents a Pearson product-moment correlation coefficient of 0.6985 with p<.0001.

The above figures show that the speakers who *tense* their short-*a* the most (before the diagnostic environment of voiced stops) also tend to both have a larger distance between their THOUGHT and LOT vowel spaces as well as show a higher percentage of THOUGHT-raising. So, while I did not find specific correlations between demographic groups of people leading the two changes, it seems that individuals are indeed acting somewhat consistently across the two phonological features. A fruitful avenue of future research would be to investigate further what the causes are of the correlation.

7.2 Authenticity and the New Yorker Identity

The "privilege" of being able to call oneself a "New Yorker" is heavily gate-kept by the residents of the five boroughs of New York City. Even among that group, there is uncertainty as to who has the authority to decide if someone is New York enough. Does it only apply if you live in Manhattan? If you are a transplant from somewhere else in the country is there a certain number of years after which you become a New Yorker, or is this status impossible to achieve unless you were born there? Growing up on Long Island, I knew from a young age to be careful saying I was New Yorker or even "from New York" in the presence of someone who lived in the actual city. In fact, I have always found it fun to teasingly point out to residents of Brooklyn and Queens that they technically do live on Long Island, because it makes them mad. It makes them mad because something about living on Long Island is inferior in their opinion. People act as if there is a stark dividing line between Queens and Nassau Counties, where the city suddenly becomes suburb, when this is not the case at all. My grandparents used to live on that dividing

line (on the Queens side) and the only effect I ever noticed was that the Nassau County snow plows would only come down half their street and stop as soon as they hit the city limits, leaving a gigantic mound of snow for my siblings and me to play in. This ideological placement of prestige is a departure from the prevailing discourse in the era after WWII, when the suburbs on Long Island started to be developed and it was a sign of prestige to be able *to leave* the city. Those people who were eager to move to the suburbs to better their lives are the parents and grandparents of many of the participants in my study, and also of myself. Of the 24 students in the study, eight of them have at least one parent who grew up in Brooklyn or Queens, and twelve of them have at least one parent who grew up on Long Island and whose grandparents had been the ones who left the city for the suburbs.

This issue of the New Yorker identity and who can lay claim to it "authentically" motivated me to ask questions in my sociolinguistic interviews about whether participants identified as Long Islanders, New Yorkers, or both. I asked questions that revealed their positive or negative feelings about both Long Island and New York City, and if they thought the two were different, both in terms of culture and in terms of accent. I encountered varying levels of success in asking such introspective questions of teenagers at such a young age, but I did glean insight into some of the students and why their linguistic behavior might be like it is. In the next section I spotlight three participants whose linguistic behavior merited further attention, Tony, Caleb, and Daniel, and briefly lay some groundwork suggesting connections between that behavior and aspects of their particular identities as Long Islanders.

7.3 Spotlight

7.3.1 Tony

Tony was the youngest mentor at 24 years old – which means that he was only six years older than the senior students in the CyberPilots. He was a student at Grumman High School not so long ago, then went away to college and immediately started mentoring the club upon returning to Long Island. Tony had the closest rapport with the students. What is noteworthy is that, despite being the youngest mentor and so close in age to the students, Tony shows a high level of NYC features in both short-*a* and THOUGHT-raising. Table 19 compares the three youngest mentors, Tony, Nate, and Mr. Price, giving their mean formant values for both THOUGHT and LOT, their Pillai scores, as well as their age, ethnicity, and short-*a* system (recall that all the mentors are male and white).

	Mean Formants for /ɔ/		**Mean Formants for /ɑ/**					
Speaker	**F1**	**F2**	**F1**	**F2**	**Pillai Score**	**Age**	**Ethnicity/ Religion**	**Short-*a* System**
Mr. Price	623	972	817	1401	0.747	30	Irish Catholic	Tragerian
Tony	679	1008	861	1278	0.660	24	Italian Catholic	Tragerian
Nate	703	1050	836	1337	0.588	31	Jewish	Transitional

Table 19: THOUGHT and LOT data for the three young mentors

Mr. Price leads in THOUGHT-raising and Pillai score, but Tony joins him in high rates of THOUGHT-raising and having a Tragerian short-*a* split, while Nate, who is seven years older than

Tony, has the lowest Pillai score and the highest mean F1 for THOUGHT, plus he has a Transitional rather than a Tragerian short-*a* system.

The main social characteristic that Mr. Price and Tony share as opposed to Nate are that they are two of the only white participants in the study who are Catholic rather than Jewish. Mr. Price is Irish and Tony is Italian. Italian is the other prominent ethnicity associated with NYC English and Long Island identity, alongside being Jewish. To investigate this possible Italian angle a little further, let's look at Chrissy, the only non-Jewish white student who is, like Tony, Italian Catholic. In Table 20 the results for both THOUGHT/LOT and short-*a* systems are listed for each of the five white female participants in the study.

	Mean Formants for /ɔ/		**Mean Formants for /ɑ/**				
Speaker	**F1**	**F2**	**F1**	**F2**	**Pillai Score**	**Age**	**Short-*a* System**
Chrissy	704	1099	841	1311	0.567	Senior	NYC-Nasal
Leah	709	1104	826	1332	0.507	Junior	NYC-Nasal
Becky	731	1147	836	1315	0.396	Freshman	Nasal Split
Chloe	753	1118	817	1289	0.264	Freshman	NYC-Nasal
Hannah	735	1105	786	1222	0.229	Junior	NYC-Nasal

Table 20: THOUGHT/LOT and Short-*a* results for the five white, female students.

Chrissy has a Pillai score of .567 and an NYC-Nasal system. Out of the white female students, Chrissy does have the highest Pillai score and the lowest mean F1 for THOUGHT at 704Hz.

These comparisons suggest that there might be something socially significant about being white and not Jewish in Antioch. Both Italian and Jewish are ethnicities that have been strongly

associated with NYCE as well as with Long Island. The associations with these groups are different, with Italian being more associated with working class and the South Shore "Strong Island" vibe described in Olivo 2013. There are only Tony and Chrissy to test this hypothesis in my data sample, but the Italian Long Island identity seems to result in higher rates of these linguistic features even if (or perhaps because of?) not living in an Italian dominant area. This is a ripe area for future research with a population that would allow further comparison of Italian and Jewish identities and linguistic behavior on Long Island.

7.3.2 Caleb

Caleb stands out as a freshman student who has some of the most "New York City" productions of both of the linguistic variables examined herein. Figure 59 below shows Caleb's THOUGHT/LOT distinction, which represents a Pillai score of .660.

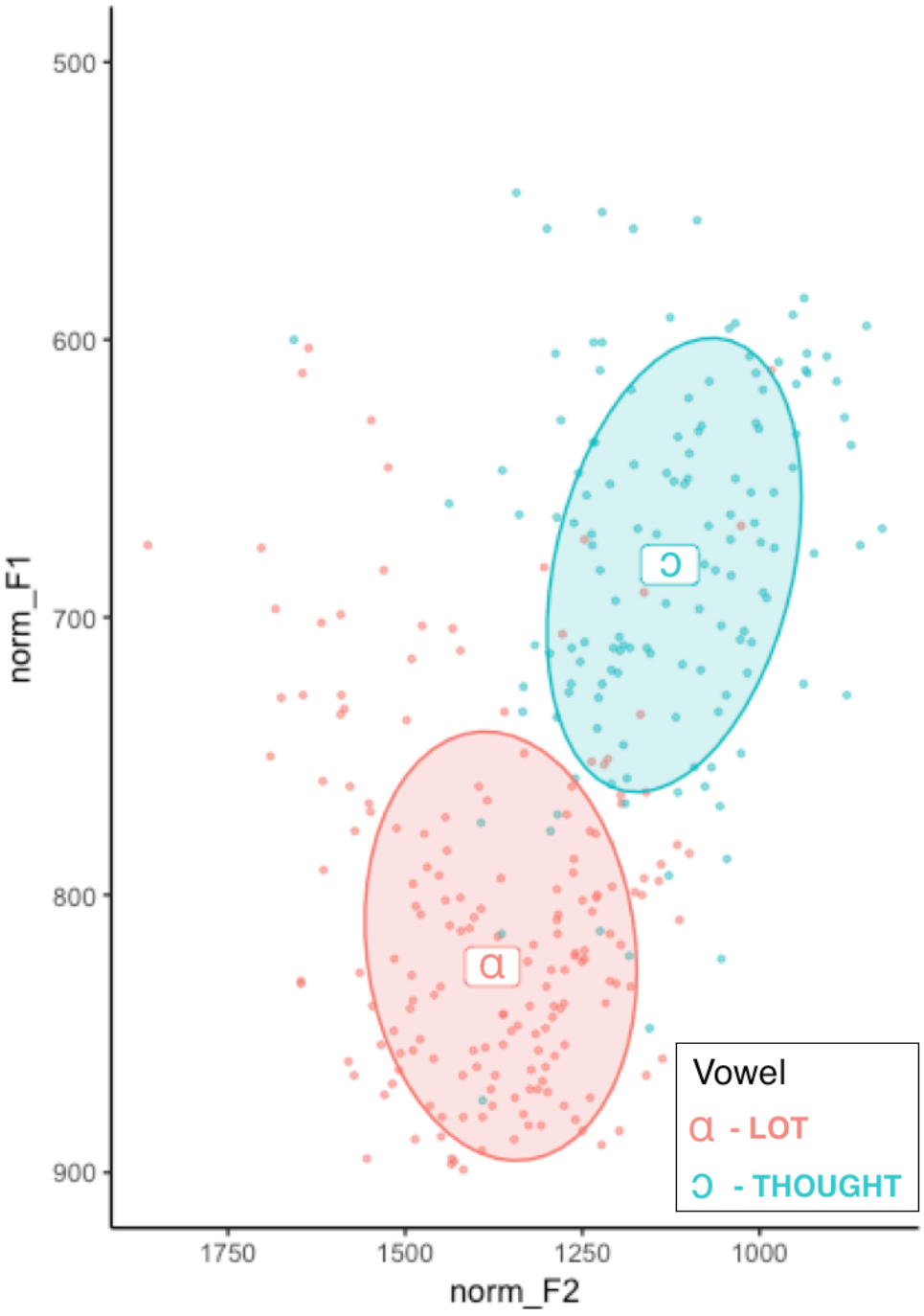


Figure 59: Vowel Plot of Caleb's THOUGHT and LOT vowels

While many students the same age or older than Caleb showed various levels of overlap between these two phonemes, the ellipses on Caleb's graph do not overlap at all. His two vowels are quite distinct, and his tokens are relatively consistent, illustrated by the small range of the ellipses. Figure 60 below illustrates Caleb's tokens of short-*a*.

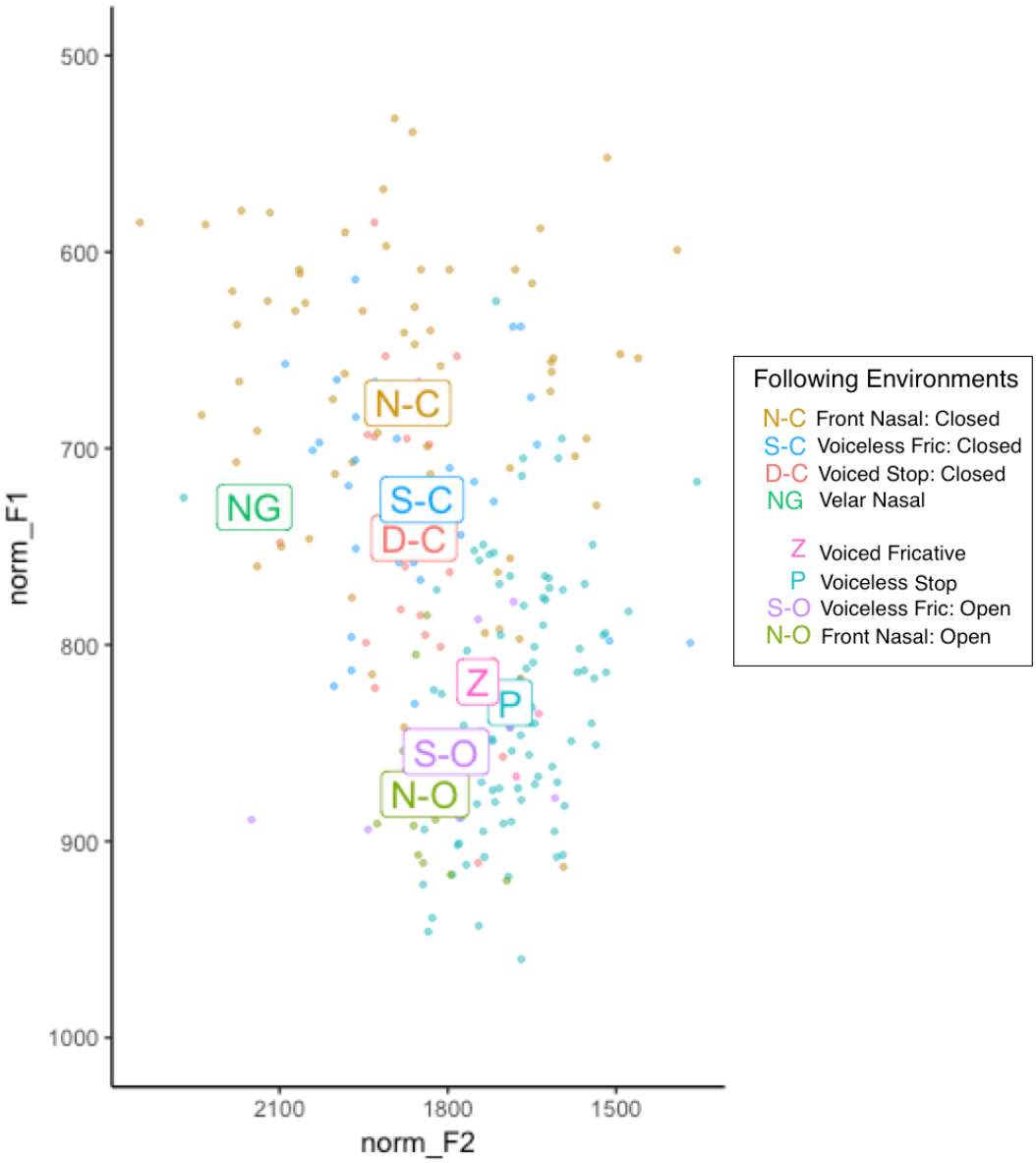


Figure 60: Vowel Plot of Caleb's tokens of Short-*a*

Caleb's plot shows a clear gap between the two variants of the vowel, with the Tragerian *tense* environments (following front nasal, voiceless fricative, and voiced stops in closed syllable) close together on one side of his split, and the Tragerian *lax* environments (following voiceless stops, voiced fricatives, and open syllables contain front nasals or voiceless fricatives) on the other.[20] Caleb's velar nasal environment is also *tense.*

A unique social characteristic of Caleb is that he has a very large family who all live in Antioch. His family is very enmeshed in the local community. Caleb specifically said that he likes Long Island and prefers it over the bustle of the city. Below is a quote by Caleb from his sociolinguistic interview, expressing his opinion on this matter in his own words.

> **Interviewer:** So, could you see yourself continuing to live on Long Island?
> **Caleb:** Yeah, I couldn't see myself going to the city. Just parking, driving's a mess, and all that in the city. Long Island's none of that. You get the same feeling because it's New York, we're New Yorkers, but you know, it doesn't come with everything else with the city.

In this quote, Caleb asserts his belief that he has an authentic claim to being a New Yorker, his living on Long Island and disliking New York City notwithstanding. He conflates the city and Long Island as one entity that is just "New York."

[20]Caleb did not produce any tokens in the D-O environment (following voiced stop in an open syllable).

7.3.3 Daniel

With regard to THOUGHT-raising, Daniel stood out as very different from the rest of the students. He patterned most closely with the older mentors. (Remember that age was a significant predictor of this trend towards the lowering of THOUGHT, so he wasn't just at the top of the pack as a senior student, but he, in fact, spoke much more like his 60-year-old mentors). I double-checked some of his highest THOUGHT tokens to make sure they were not mis-measurements, but they are genuine. Figure 61 below shows the stark distinction between Daniel's THOUGHT and LOT vowels, which represents a Pillai score of .811 (the second highest out of students and mentors alike, after only Mr. Spector). Figure 62 reproduces Figure 54, which illustrates just how much higher Daniel's THOUGHT tokens are than his peers. Daniel's *outliers* for F1 of THOUGHT are even at the 700Hz line.

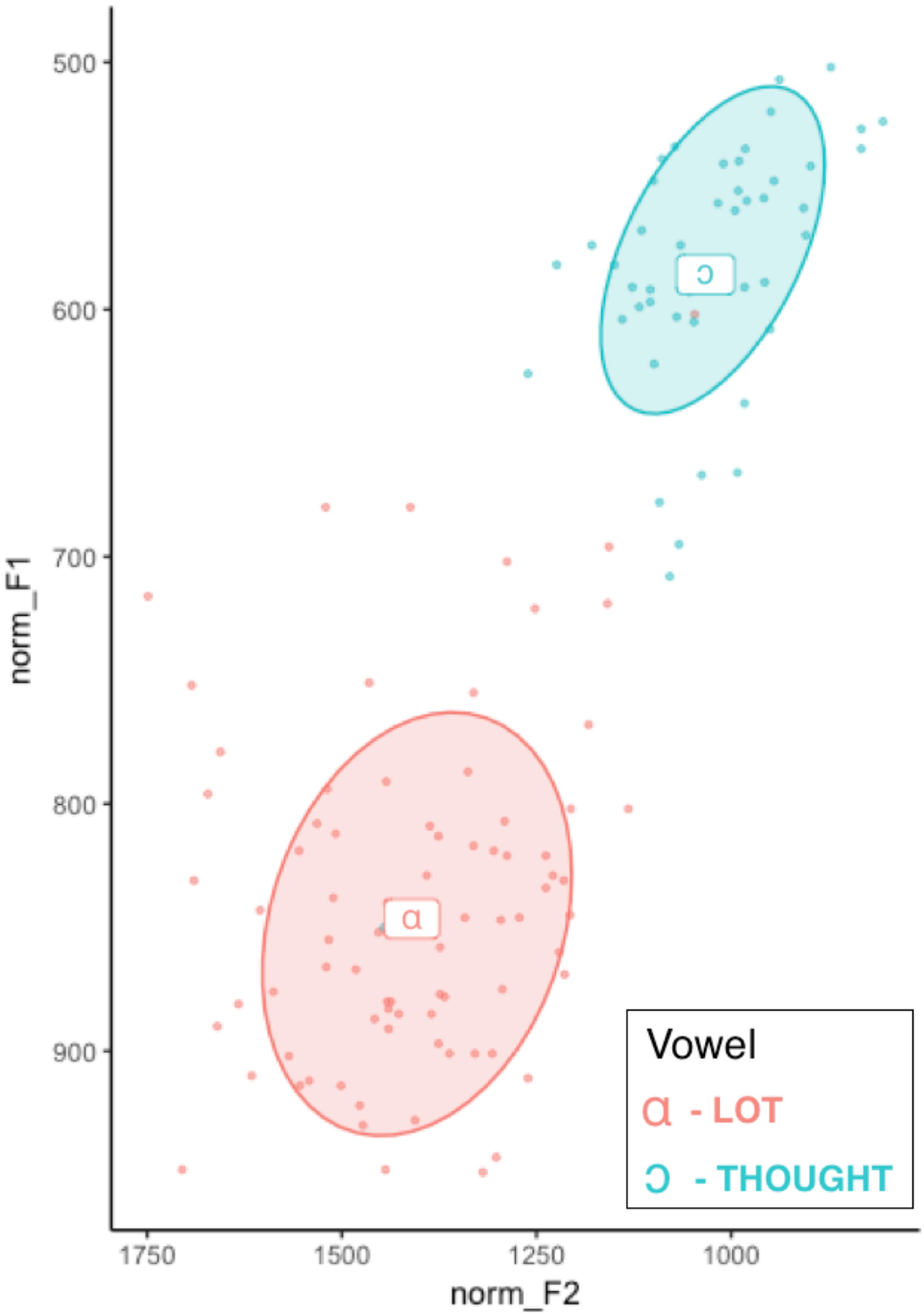


Figure 61: Daniel's THOUGHT/LOT Distinction, Pillai score of .811

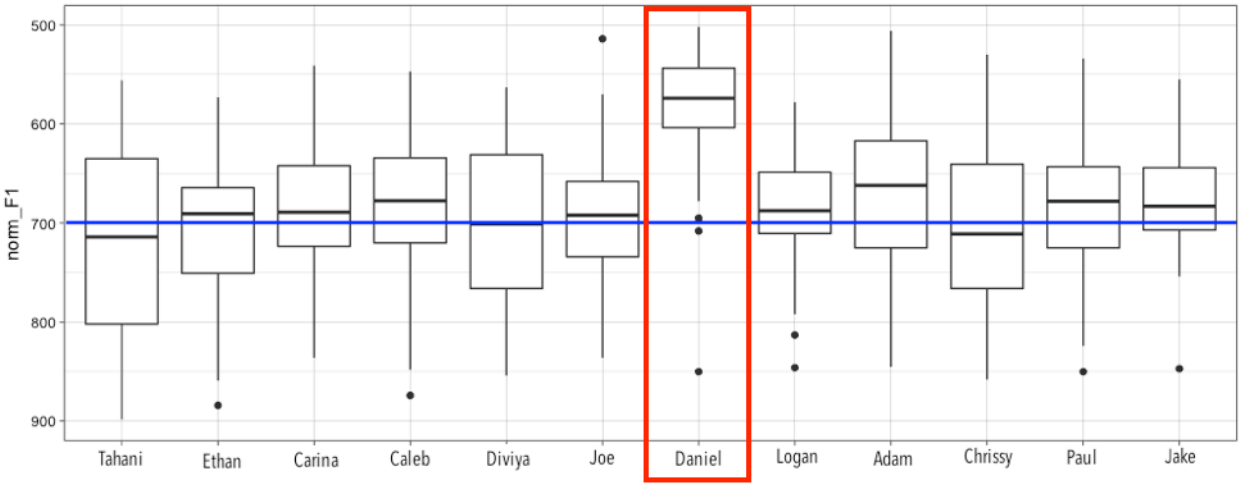


Figure 62: Boxplots of the top twelve student raisers of THOUGHT, with Daniel highlighted

So then, what about Daniel's social characteristics and attitudes about Long Island might contribute to his extreme showing of NYC features? Daniel has one parent who grew up in Queens and one parent who grew up on the South Shore of Nassau County. He is very locally oriented in his own attitudes about living on Long Island. When I interviewed him, he was in the process of choosing which college to attend, and all of his choices were on Long Island. He planned to stay living with his parents and commute, which is not extremely common and is lightly stigmatized in Antioch. The only travelling he had ever done was to Florida to see his grandparents, and sometimes to upstate New York, he said. He has never travelled elsewhere in the United States or out of the country. When asked if he goes to the city at all, he just said "No, not really," whereas most students had at least some stories about visiting the city to see shows or visit family. Daniel's social activities outside of robotics all revolve around his Jewish youth groups. He spoke about attending Hebrew High School (which is optional Hebrew school after

one has their bar or bat mitzvah; the majority of Antioch Jewish students end their studies after that event) and being very involved with USY (United Synagogue Youth).

I asked Daniel if he thought his mother had a New York City accent, since she grew up in Queens, and he vehemently said, "No, not at all." I asked about his grandparents, who also both grew up in Queens, and I received another flat-out "no." I asked Daniel if he is aware that there *is* a NYC accent and he said yes. I asked if Long Island has its own accent and he said "No, we all sound the same," referring to Long Island and New York City as one "we" that has no differentiation in accent.

Daniel is very locally oriented to Long Island, has no plans to move away, and also has, by far, the strongest New York accent of any of the students. Like Caleb, he stands out as having particularly high rates of the NYC features analyzed here, as well as being one of the most locally Long-Island oriented of the students. This is the line of inquiry that I would most like to continue in future work. As mentioned above, I had difficulty getting as much reflection from the students about their identities and relationships with their geographical surroundings as I had hoped, but there are other avenues and methodologies that could be used to more deeply investigate this relationship between Long Island local identity and high rates of use of NYCE features.

7.4 Conclusion

In this dissertation I have set out to document and investigate the state of two phonetic features of New York City English as they are used now in Nassau County. The study centered

on teenagers because this age range is the adolescent peak of language change, and these young speakers would be the *least* likely to display these features if they were indeed diminishing as has been found in Manhattan. Looking at teenagers and finding these features provides strong support for their continued use in Nassau County.

The speakers whom I highlighted in this chapter each relate to the main social questions I mean to raise with this research: the issue of the "New Yorker" identity, who can claim it authentically, and how it currently corresponds (or doesn't to the linguistic markers that have traditionally been indexically associated with it. Tony, whose role is as a mentor but whose age is closer to that of the students, lets us start to tease apart how much of the difference between mentors and students (which were significant differences in all measures) is due to actual age, rather than possibly other identity factors such as ethnic groups other than the majority Jewish population of Antioch.

I highlight Caleb and Daniel to get at core issue of the Long Island identity and how is it intertwined with that of New Yorker. Both of these students had higher rates of THOUGHT-raising and more conserved Tragerian splits than the rest of the students. They are both very locally oriented, with family living close together in Nassau County, and they both expressed negative sentiments about New York City and a desire to continue living on Long Island when they grow up.

I have shown in this dissertation that two features of New York City English, the Tragerian short-*a* split and THOUGHT-raising, are still present among young speakers in Nassau County, and that the young people who have the highest rates of these features are actually the ones with the most local attitudes and connections, rather than those with more exposure to New

York City. These results may suggest that rather than following the same trajectory as Manhattan and losing these features altogether among young speakers on Long Island, the same features might be in the midst of shifting their sociolinguistic meaning and becoming associated with an emerging “Nassau County English.”

Appendices

Appendix A: Example Interview Modules

I. Demographics/family
   a. When were you born? (Oh, so you're a….junior/senior?)
   b. Where were you born? Did you grow up here?
   c. Have you lived anywhere else? For how long?
   d. Tell me about your family:
      i. Where are your parents from?
         1. Where do they live now?
         2. What do they do?
         3. What languages do they speak?
      ii. If you have brothers or sisters:
         1. Where were they born? Where do they live now?
         2. Are you close to your siblings?
         3. Do they go to the same school? Do you like or dislike that?
      iii. Do you have a large extended family?
         1. Where do they live?
         2. Are you close/do you see them often?
         3. What is one of you most memorable moments with your family or extended family?
      iv. Does most of your family live on Long Island or in the New York City area?
   e. What languages were spoken at home when you were growing up?

II. Robotics
   a. How long have you been involved in robotics?
   b. What first interested you in robotics and/or engineering?
   c. What do you specialize in on the team?
      i. Do you think there is a difference between the types of people who work on different aspects?
   d. Do other people in school see robotics kids as "geeks?" How do you see yourself?
   e. What is your favorite thing about being part of the robotics team?
   f. Have you been to any of the competitions?
      i. What was your role at the competition? Were you a driver or human player? Did you present to the judges?
      ii. Did anything go wrong? How did the team deal with it?
   g. Do you plan to do engineering or robotics in college? As a job?

III. Social Media
    a. What forms of social media do you use regularly?
        i. ex: Instant Message (Gchat, Facebook), text messages, Twitter, Whatsapp, Snapchat, Instagram, etc.
    b. Which would you say you use the most? How often do you use them each day?
    c. What forms of social media do your friends use? Are they the same as yours?
    d. Do you think older people "get" the way you use social media? Would they understand what you're saying?
    e. Do you think you speak differently when writing online than you do when speaking?
    f. What about in different activities like text messaging vs. Facebook – do you think you talk differently depending on where you're posting?
    g. Do you talk to your parents online? With what mediums?

IV. Long Island/Language
    a. Do you like living on Long Island?
        i. Would you rather live somewhere else? Where?
    b. Where do high schoolers on Long Island go to hang out? What do you do?
    c. Does Long Island have its own accent?
    d. What have you heard about the way Long Islanders talk?
        i. Do you think there is any truth to it?
        ii. Do you think it is different from how people in New York City talk?
    e. Do older people on Long Island speak differently than younger people?
    f. Do people speak differently depending on which part of the island they are from?
    g. When you go other places, do people ask you where your accent is from?
    h. How do people in other parts of the U.S. sound

Appendix B: Reading Passage

Based on an interview with comedian Louis CK

My first time on stage was horrible. I found out that there was an open mic night at a club where I grew up in Boston, called Stitches. I was actually 17, almost 18, years old, and I thought, jeez, you can just go on stage. You put your name in a hat, and you get called on stage. You can be anybody. Although daunting, that was so exciting to me.

So I went to this club Stitches, which is very grown-up. The manager didn't even want to let me in because I was underage and they have a liquor license. And then I went on stage, wearing this horrible orange shirt, the height of 80s fashion. I did about two minutes because I didn't have enough material. I just ran out, I sputtered. My whole throat constricted and I heard this roaring in my ears. My eyes were watering. My heart was pounding, and I fought to control myself. I couldn't think straight. And all these adults just sat there and looked at me like I was an idiot. I coughed, and then I just walked off stage to crash in the bathroom.

Well, I didn't do it again for a little while after that. At the time, I was working at a video rental store. It just so happened that Paul Hodgeman, who was a very big star in the Boston comedy scene back then, was a customer at the store. I told him that I had done an open mic, and he said, "Oh, well, you've got to come on my show." He had a show that was a huge thing. He had this show where they would pack the room and put on all the best comics in Boston. And he said, come be on my show. And I said, "well, no, I need to practice and, like, you know, do more shows to get good enough." And he said, "no, I won't let you go on my show unless you go on now," very authoritatively. He said, "It's interesting because you don't know what you're like, it's comedy tragedy." He kind of wanted me to go on and bomb.

They were shocked at how bad I was - just shocked. I got off stage and Paul wouldn't look me in the eye. Nobody would. I was exhausted, and it was just the most pulverizing humiliation. It was much worse than the first time because the first time was an open mic night. The premise of the show is that most of us don't know what we're doing. But this was a professional comedy show, and I went on it and just flailed. I mean, it was a nightmare.

But it turned out to be a great thing. Because it gave me a very realistic, shocking picture of what I was facing. That show that I did and that audience and that night - that is still the terrain that I work on today. I think if you're just looking for easy ways, or you're looking only for victories throughout life, I don't think you're really getting much out of it. I think if you really know how hard stuff is and despite that, you extract some tools out of yourself, it's better for you. And the next time I went on stage, I was very wary.

Appendix C: Word List

tab	cabin	botch	apple	mauve	coffee
zed	thought	manage	math	and	gaunt
soon	pass	bat	hash	poverty	talk
song	caddy	annex	cash	fawn	passive
hock	niece	human	thieve	humid	bad
cab	pawn	pet	peas	knees	awful
hanging	conflict	humor	father	boom	madden
gong	tom	set	huge	sat	pot
Kong	get	brawn	bought	leaf	awed
ham	bead	moss	got	hog	cats
hodge	happen	hospital	dawn	caught	cabs
cause	daunt	sang	sauce	hang	debt
tune	shop	ask	boot	cough	I can
back	tin can	met	have	laundry	deck
half	sob	piece	man	sad	dam
bath	bed	moot	thief	passion	tap
jazz	gaffe	moon	gas	coffin	bag
hawk	chalk	pause	hash	god	jazzy
moose	calm	addict	gosh	mom	tong
bet	bomb	Aztec	palm	savvy	prove
net	dodge	broad	leave	dog	fraud
cod	pong	notch	goth	mannin	sack
proof	beat	cap	hanger	g	toss
after	tooth	sag	pang	damage	
avenue	con	gossip	pod	cot	

References

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge, UK ; New York: Cambridge University Press.

Babbitt, E. H. (1896). The English of the lower classes in New York City and vicinity. *Dialect Notes*, *1*, 457–464.

Bailey, G. (2002). Real and apparent time. In *The Handbook of Language Variation and Change*. Malden, MA: Blackwell Publishing.

Bakht, M. (2010). *Lexical Variation and the Negotiation of Linguistic Style in a Long Island Middle School* (PhD Dissertation). New York University.

Becker, K. (2010). *Regional Dialect Features on the Lower East Side of New York City: Sociophonetics, Ethnicity, and Identity.* (PhD Dissertation). New York University.

Becker, K., & Wong, A. W. (2010). The Short-a System of New York City English: An Update. *University of Pennsylvania Working Papers in Linguistics*, *15*(2), 11–20.

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer (Version 6.0.35). Retrieved from http://www.praat.org/

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Charter of the Dutch West India Company. (1621). Retrieved December 1, 2018, from http://avalon.law.yale.edu/17th_century/westind.asp

Coggshall, E. L. (2017). *Short-a in the Sixth Borough: A Sociophonetic Analysis of a Complex Phonological System in Jersey City* (PhD Dissertation). New York University.

Coggshall, E. L., & Becker, K. (2010). The vowel phonologies of African American and White New York City residents. In *African American English Speakers and their Participation in*

Local Sound Changes: A Comparative Study. (pp. 101–128). Durham, N.C.: Duke University Press.

Cohen, P. (1970). *The Tensing and Raising of Short (a) in the Metropolitan Area of New York City.* (MA Thesis). Columbia University.

Dinkin, A. J. (2011). Weakening resistance: Progress toward the low back merger in New York State. *Language Variation and Change*, *23*(3), 315–345. https://doi.org/10.1017/S0954394511000147

Frank, Y. H. (1948). *The Speech of New York City* (PhD Dissertation). University of Michigan.

Gouskova, M. (2004). Relational hierarchies in Optimality Theory: the case of syllable contact. *Phonology*, *21*(2), 201–250. https://doi.org/10.1017/S095267570400020X

Hall-Lew, L. (2010). Improved representation of variance in measures of vowel merger. *Proceedings of Meetings on Acoustics*, *9*(1), 060002. https://doi.org/10.1121/1.3460625

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Hubbell, A. F. (1950). *The pronunciation of English in New York City.* New York: King's Crown Press.

Jaggers, Z. S. (2018). Evidence and characterization of a glide-vowel distinction in American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 3. https://doi.org/10.5334/labphon.36

Johnstone, B. (2016). Enregisterment: How linguistic items become linked with ways of speaking. *Language and Linguistics Compass*, *10*(11), 632–643. https://doi.org/10.1111/lnc3.12210

Kahn, D. (1976). *Syllable-based generalizations in English phonology.* (Thesis). Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/16397

Kenyon, J. S. (1924). *American Pronunciation*. Ann Arbor, MI: George Wahr Publishing Company.

Kurath, H. (1949). *A Word geographie of the Eastern United States* (4. print). Ann Arbor, Mich: Univ. of Michigan Pr.

Kurath, H., & McDavid, R. I. (1961). *The pronunciation of English in the Atlantic States: based upon the collections of the linguistic atlas of the Eastern United States*. Ann Arbor, MI: University of Michigan Press.

Labov, W. (1966). *The social stratification of English in New York City* (2nd ed). Cambridge [UK] ; New York: Cambridge University Press.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: Univ. of Pennsylvania Press.

Labov, W. (1981). Resolving the Neogrammarian Controversy. *Language*, *57*(2), 267–308. https://doi.org/10.2307/413692

Labov, W. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change.* Berlin: Walter de Gruyter. Retrieved from http://public.eblib.com/choice/publicfullrecord.aspx?p=453906

Labov, W. (2007). Transmission and Diffusion. *Language*, *83*(2), 344–387.

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language*, *89*(1), 30–65. https://doi.org/10.1353/lan.2013.0015

Labov, W., Yaeger, M., & Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress.* Philadelphia: US Regional Survey.

Laferriere, M. (1979). Ethnicity in Phonological Variation and Change. *Language*, *55*(3), 603–617. https://doi.org/10.2307/413319

Lambert, B. (1997, December 28). At 50, Levittown Contends With Its Legacy of Bias. *The New York Times*. Retrieved from https://www.nytimes.com/1997/12/28/nyregion/at-50-levittown-contends-with-its-legacy-of-bias.html

Newlin-Łukowicz, L. (2016). *Ethnicity, L1 Interference, and Sound Change in New York City* (PhD Dissertation). New York University.

Newman, M. (2010). Focusing, implicational scaling, and the dialect status of New York Latino English1. *Journal of Sociolinguistics*, *14*(2), 207–239. https://doi.org/10.1111/j.1467-9841.2010.00441.x

Newman, M. (2011). The NYC short-a system: Continuity and change over generations in a complex phonological system. Presented at the NWAV 40, Georgetown University.

Newman, M. (2014). *New York City English*. Boston: De Gruyter.

Nycz, J., & Hall-Lew, L. (2013). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*, *20*(1), 060008. https://doi.org/10.1121/1.4894063

Office, A. C. S. (n.d.). 2016 Data Profiles. Retrieved December 1, 2018, from https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2016/

Olivo, A. M. (2013). *The Strong Island Sound: Sociolinguistic Evidence for Emerging American Ethnicities* (PhD Dissertation). Rice Universtiy.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency Effects and the Emergence of Lexical Structure* (pp. 137–157). Amsterdam: John Benjamins.

Press, L. I. (2017, September 9). Top 10 Differences Between the North and South Shores of Long Island. Retrieved December 1, 2018, from

https://www.longislandpress.com/2017/09/09/top-10-differences-between-the-north-and-south-shores-of-long-island/

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). *FAVE (Forced Alignment and Vowel Extration) Program Suite*. Retrieved from fave.ling.upenn.edu

Seyfried, V. F. (1984). *The Long Island Rail Road: a comprehensive history*. *The age of electrification, 1901-1916*. (Vol. 7). Garden City, Long Island: Vincent F. Seyfried.

Slomanson, P., & Newman, M. (2004). Peer group identification and variation in New York Latino English laterals. *English World-Wide*, *25*(2), 199–216.

Sneller, B. (2018). *Mechanisms of Phonological Change* (PhD Dissertation). University of Pennsylvania.

The Gallup Organization. (2016). Five Key Findings on Religion in the U.S. Retrieved December 1, 2018, from https://news.gallup.com/poll/200186/five-key-findings-religion.aspx

Thomas, E. (2001). An Acoustic Analysis of Vowel Variation in New World English. Durham, N.C.: Duke University Press.

Trager, G. L. (1930). The pronunciation of "short a" in American Standard English. *American Speech*, *5*(5), 396–400.

Trager, G. L. (1934). What conditions limit variants of a phoneme? *American Speech*, *9*(4), 313–315.

Trager, G. L. (1940). One Phonemic Entity Becomes Two: The Case of "Short A." *American Speech*, *15*(3), 255–258. https://doi.org/10.2307/486966

U.S. Census Bureau QuickFacts: Nassau County, New York; UNITED STATES. (n.d.). Retrieved November 29, 2018, from https://www.census.gov/quickfacts/fact/table/nassaucountynewyork,US/PST045217

Wong, A. W. (2007). Two vernacular features in the English of four American-born Chinese. *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 217–230.

Wong, A. W. (2012). The lowering of raised-THOUGHT and the low-back distinction in New York City: Evidence from Chinese Americans. *University of Pennsylvania Working Papers in Linguistics*, *18*(2), 157–166.

Wong, A. W. (2015). *Diverse Linguistic Resources and Multidimensional Identities: A Study of the Linguistic and Identity Repertoires of Second Generation Chinese Americans in New York City* (PhD Dissertation). New York University.

Wong, A. W., & Hall-Lew, L. (2014). Regional variability and ethnic identity: Chinese Americans in New York City and San Francisco. *Language & Communication*, *35*, 27–42. https://doi.org/10.1016/j.langcom.2013.11.003