Sounding Guilty: Criminality and Black Racialized Speech

by

Dominique Branson

Bachelor of Arts, Cornell University, 2014

Master of Arts, University of Pittsburgh, 2022

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Dominique Branson

It was defended on

June 30, 2023

and approved by

Dr. Amanda Godley, Professor, Department of Teaching, Learning, and Leading

Jasmine Gonzales Rose, JD, Professor of Law, Center for Antiracist Research

Dr. Junia Howell, Visiting Assistant Professor, Department of Sociology

Dissertation Director: Dr. Shelome Gooden, Professor, Department of Linguistics

Dissertation Co-Director: Dr. Marta Ortega-Llebaria, Associate Professor, Department of Linguistics

Sounding Guilty: Criminality and Black Racialized Speech

Dominique Branson, PhD

University of Pittsburgh, 2023

A widely held belief in the United States asserts that race is seen but not heard (Gonzales Rose, 2018). According to this, racialization and racism are not enacted through hearing alone. However, a large and consistent body of sociolinguistic research has shown that listeners do indeed racialize disembodied voices, and often categorize speakers' racial identities accurately even when given no additional external cues to decipher this information. (Haley, 1990; Purnell, Idsardi & Baugh, 1999; Thomas & Reaser, 2004; Weissler, 2021). Linguists have also shown that listeners may negatively assess racialized speakers and unfairly discriminate against them, using speakers' speech practices as justification for their evaluations and actions (Flores & Rosa, 2015; Rosa & Flores, 2017). An emerging body of research on Blackness and language in legal settings, for example, has shown that court outcomes may be impacted by listeners' negative judgments of Black speakers who use African American Vernacular English (AAVE) when testifying in the courtroom (Corley [Branson], 2014; Kurinec & Weaver III, 2019; Jones et al., 2019; Rickford & King, 2016). The present study builds upon this work by investigating whether listeners' perception of speakers' racial identities correlates with listeners' perception of criminality.

I designed three surveys to investigate whether listeners associate voices they categorize as "sounding Black"—what I call *Black racialized speech*—with crime relevant images. I found a small positive correlation ($r = 0.213$) between these variables among listeners who heard Black speakers. While not statistically significant, this suggests that listeners may associate Black racialized speech with criminality, but only when the speakers they are listening to self-identify as

Black—and this is despite the fact that listeners do not know speakers' racial identities nor have additional external cues to decipher this information.

Together, the findings highlight the critical role of listeners' perception in racialization and criminalization processes and emphasize that moving towards racial justice using linguistics necessitates greater emphasis on interrogating listeners' perception. Finally, I assert that Black racialized speech is a useful heuristic tool to begin to interrogate the ways in which non-production-based factors influence listeners' perception of speakers/speech in racialization and criminalization processes.

Table of Contents

List of Tables

List of Figures

Preface

My dissertation expands upon *and* dismantles my very early research *Sounding Guilty: African American English and Racial Inequality in the Courtroom* (Corley [Branson], 2014), from which the present study derives part of its name. It expands upon my earlier work because my reason for doing it has not changed. Ten years ago, I did not want a Black person to go to jail because someone else could not understand what they were saying—and I still don't. However, early on, I struggled to translate this desire into impactful research. Despite this, during the summer before my senior year of undergrad, I seized an opportunity to conduct independent fieldwork in Detroit, Michigan through a Mellon Mays undergraduate fellowship. I told every person I met that summer about my research interest, and one day, someone asked me if I had seen the witness's testimony in the Trayvon Martin case. She said, you *need* to watch it. I wish I could remember who said this to me, because I would love to thank her and tell her about the incredible journey she sent me on with that single comment. She was talking about *Rachel Jeantel*.

Rachel Jeantel was a friend of Trayvon Martin's and the last person to hear him alive before he was killed by George Zimmerman in 2012. She was the prosecution's star witness, and for six hours on the stand she described her final phone call with Martin, all the while speaking African American Vernacular English (AAVE). Yet, the jury disregarded her testimony and found George Zimmerman *not guilty* (Rickford & King, 2016). I concluded, like other sociolinguists, that the jurors dismissed Jeantel's testimony because of their negative attitudes towards AAVE. Had she spoken standardized American English (sdAE), perhaps there would have been a different outcome.

Like a good social scientist-in-training, I tested this hypothesis empirically when I returned to Cornell in the fall. I developed a matched-guise experiment using Jeantel's AAVE testimony and an sdAE testimony I created based on the AAVE one. I distributed the experiment to 24 of my peers who listened to and assessed one or the other testimony. I found that the listeners did judge Jeantel's testimony differently when they heard it in AAVE compared to sdAE. So, I concluded that had she spoken sdAE during her testimony, maybe Zimmerman would now be behind bars (Corley [Branson], 2014). Today, not only do I feel troubled about putting anyone behind bars, but I no longer believe that the outcome of the case had much to do with the way Jeantel spoke during her testimony. Rather, I think the jurors' dismissal of her testimony was likely based on their own biased judgments of Jeantel and her racial identity. Thus, the present study is also a dismantling of my earlier work.

In alignment with critical scholars and raciolinguists, in this dissertation, I argue that the practices and processes that criminalize Black speakers and their language are less the result Black speakers' physical production of speech and more the product of listeners' own biased listening practices (Alim et al., 2016; Flores & Rosa, 2015; Rosa & Flores, 2017). In fact, my dissertation shows that, while there is a positive correlation between Black speakers' use of some prominent AAVE features and listeners thinking speakers "sound Black," this does not tell the full story. Instead, I show that there is a need to demonstrate how listeners' racialization and criminalization of Black speakers are linked to listeners' own biases and institutionalized racism. I therefore urge sociolinguists and other researchers to centralize the ways in which listeners' practices are implicated in processes that criminalize Black speakers, without first trying to show how these processes are linked to Black speakers' language use. Importantly, I also advocate for addressing the structures and institutions that perpetuate the criminalization of Black people in America and

thereby shape the listening practices that link Black speakers to criminality in U.S. society. *Whoever has ears, let them hear* (Matthew 13:9, New International Version).

I received support from so many people to get to this place in my doctoral career. I would like to thank my committee for their time and dedication to see this work through. I would especially like to acknowledge my colleagues and students at the University of Illinois at Chicago—thank you for helping reshape how I think about justice. To Gabbie and Lorenzo, thank you for being my research assistants and supporting every wild idea I had about this research. I have so many friends and family members to thank as well. Most of all, thank you Kam and Tevin for the constant encouragement, prayers, motivation, and support. I am beyond grateful.

1.0 Introduction: *Whoever Has Ears…*

"Water for sale! $2.00 a bottle!"
Jordan Rodgers, 8 years old

"All I heard was the yelling. How could it possibly be racism?"
Alison "Permit Patty" Ettel

In 2018, an eight-year-old Black girl, Jordan Rodgers, advertised the sale of $2.00 water bottles in front of her apartment complex. She repeated, "Water for sale! $2.00 a bottle!" to grab the attention of potential customers nearby (Jefferson-Jones & Henderson, 2020). After a while, another resident—a middle-aged White woman named Alison Ettel—approached the girl and then called the police on her allegedly for, "illegally selling water without a permit" (Jefferson-Jones & Henderson, 2020, p. 876). Rodgers' mother, Erin Austin, recorded the encounter on her phone, posted the video on social media, and it later went viral. Social media users accused Ettel of being racist, nicknaming her "Permit Patty" (Jefferson-Jones & Henderson, 2020). Ettel countered, saying her actions were not racially motivated. She said, "All I heard was the yelling. How could it possibly be racism?" (Jefferson-Jones & Henderson, 2020, p. 876). Great question, Permit Patty.

1.1 Overview

The question, *how could it possibly be racism*, illustrates a widely held belief in the United States that race is seen, but not heard (Baugh, 2005; Gonzales Rose, 2018; Weissler, 2021[1]).

[1] These articles highlight the fallacy of this ideology.

According to this logic, then, racism, racialization–and in the case above, criminalization–cannot be enacted solely through hearing, or the act of listening. Yet, a large and consistent body of sociolinguistic research has demonstrated that listeners do indeed racialize voices, and they do so fairly accurately, or in ways that match speakers' self-identified race (Baugh, 1996; Haley, 1990; Purnell, Idsardi & Baugh, 1999; Thomas & Reaser, 2004; Weissler, 2021). In addition, some sociolinguists have conducted empirical research showing that disembodied Black-sounding voices may be criminalized in mock courtroom experiments (Corley [Branson], 2014; Kurinec & Weaver III, 2019; Jones et al., 2019). Taken together, this body of research has begun to reveal that listening practices may be racist. To add to research on Blackness, language and criminality, the present study focuses on two under-investigated areas, detailed just below.

1.2 Under-Investigated Areas in Blackness, Language, and Criminality Research

First, much of the recent research on the relationship between Blackness, language, and criminality has centralized courtroom contexts—especially defendant and witness testimonies (Corley [Branson], 2014; Kurinec & Weaver III, 2019; Jones et al., 2019; Rickford & King, 2016). Historically and presently, however, the criminalization of Black Americans has occurred all throughout U.S. society and not just within courts. This is not only evidenced by the alarming racial disparities at each juncture in the criminal legal system–profiling, arrests, guilty verdicts, sentencing, and incarceration, to name a few (Alexander, 2020; Hinton & Cook, 2021; Johnson & Lee, 2013; Kovera, 2019; Pager, 2003; Sommers & Marotta, 2014; Yang, 2015)–but also by Permit Patty's criminalization of an eight-year-old Black girl simply *#LivingWhileBlack* (Thurston, 2019). Therefore, this dissertation aims to increase the ecological validity of research on

Blackness, language, and criminality by investigating how linguistic practices contribute to Black criminality in multiple contexts throughout U.S. society.

Second, the current research has centralized Black speakers' production of African American Vernacular English (AAVE). The work contends that listeners' interpretation and judgments of *AAVE* influence how they racialize and, thus, criminalize Black speakers (Corley [Branson], 2014; Kurinec & Weaver III, 2019; Jones et al., 2019; Rickford & King, 2016). Put another way, it asserts that Black speakers' production and listeners' assessments of AAVE mediate listeners' judgments of Black speakers. Yet, sociolinguists have not always been able to successfully isolate the AAVE features that differentiate listeners' racialized categorizations and assessments of speech (i.e., listeners' judgments that speech "sounds Black" or "sounds White") (e.g., Gaither et al., 2015; Thomas & Reaser, 2004). This suggests that speaking AAVE may be one way to "sound Black," but is not the only way nor the only factor shaping how, when, and why listeners think speakers "sound Black" or criminalize "Black-sounding" speech.

Moreover, research in raciolinguistics and psycholinguistics has theorized and empirically demonstrated that listeners' assessments of speakers are largely (and, at times, solely) based on their own listening practices—practices which are themselves shaped by listeners' experiences as well as institutionalized racism, but not necessarily Black speakers' production of speech, including AAVE (e.g., Flores & Rosa, 2015; Rosa & Flores, 2017; Rubin, 1992). In these studies, listeners had additional (e.g., visual) cues to decipher speakers' racial identities. Nevertheless, the work highlights that even for research, like the present study, that focuses on speech alone, sociolinguistics needs a listener-centered approach that is not mediated by or explained solely using Black speakers' vocal production of AAVE.

The present study bridges both gaps—the centralization of courts and AAVE in this body of research—through an empirical investigation of the relationship between Blackness, language, and criminality in U.S. society.

1.3 Research Objectives and Question

To address the first gap, I examine the ways in which listeners' practices criminalize Black speakers in U.S. society, without centralizing courts (and court testimonies). I do this by historicizing the present study (Chapter 2) and utilizing a memorization experiment to test listeners' association of "Black-sounding" speech with five different crime relevant images to represent "criminality" (Chapter 7). Though some of the crime relevant images (e.g., gavel) represent criminality/criminalization within the courtroom, others more broadly depict criminality, such as in other areas of the criminal legal system (e.g., police officer) and society in general (e.g., handgun). Thus, the study does not centralize criminalization of "Black-sounding" voices in the courtroom.

To address the second gap, I interrogate listeners' perception of Black speakers' racial identities and criminality directly, without utilizing AAVE as an intermediary, explanatory variable. I do this by deriving and implementing a listener-centered, perception-based heuristic measure, which I call *Black racialized speech* (Chapter 4). Black racialized speech represents the likelihood a speaker is categorized as "sounding Black" by listeners, and I use the measure to evaluate whether listeners' perception of speakers' racial identities (Chapter 4) correlates with listeners' perception of speakers' criminality (Chapter 7). Black racialized speech is beneficial to this research because it encompasses many diverse ways to "sound Black" that are not limited to

AAVE. In fact, some scholars have argued that AAVE primarily represents the speech of just one subgroup of Black speakers (King, 2020; Smitherman, 1988). Comparatively, Black racialized speech takes into account that Black speakers may "sound Black" to listeners even without speaking AAVE (e.g., Rubin, 1992). Yet, while this is an important benefit, it is not the primary motivation for using Black racialized speech instead of AAVE. African American Language (AAL), for example, represents language used by African Americans in communities throughout the United States, and could easily be used instead of AAVE in this research. However, AAL is also based on Black speakers' production of speech and the goal of this study is to centralize listeners' *perception* and deemphasize speakers' production. As such, the primary purpose for deriving and implementing Black racialized speech is its capacity to directly capture listeners' *perception* of speakers/speech. Specifically, it captures whether they perceive speakers/speech as "sounding Black," and then I test whether these perceptions are associated with criminality.

Based on these aims, this dissertation addresses the core question, ***Do U.S. listeners associate Black racialized speech with criminality*****?** Addressing this question using a listener-centered approach has important implications for the field of sociolinguistics as well as broader U.S. society.

1.4 Significance of the Present Study to Sociolinguistics and Society

Since research on language in Black American communities first began, sociolinguists have been motivated to address racial disparities and promote racial justice through their work (Baugh, 1979; Fasold, 1972; King, 2020; Labov, 1966; Wolfram, 1969). However, like researchers in many social science disciplines, at times, their work unintentionally reified the very hierarchies

they sought to rectify (for discussion on this see Bucholtz, 2003; King, 2020, Wolfram, 2007). Addressing the field's challenging history with race (research), in 2020, Charity Hudley, Mallinson, and Bucholtz put forth that, to advance racial justice in the field, linguists need to reconsider their approach to the study of race and racism. They explained that this can be accomplished by drawing from work in non-linguistic disciplines that focus on race, racialization, and racism to inform (socio)linguistics' conceptualization of race. In addition, they highlight that the field must make efforts to value the unique contributions of scholars of color to linguistics research (Charity Hudley, Mallinson & Bucholtz, 2020).

The present study attempts to incorporate these strategies to promote racial justice by drawing from psycholinguistics and raciolinguistics as well as critical race studies, critical criminology, psychology, and sociology to illuminate the critical need to focus on listeners' perception in linguistic research on race, racialization, and criminalization. In doing this work, I also align myself with other junior scholars of color like Wright (2017; 2019), Weissler (2022), and King (2020; 2021) who are merging disciplinary boundaries, methodologies, and theoretical frameworks to reshape how linguists engage with race in their research. They also centralize their positionality as scholars of color and underscore the unique perspectives they add to the study of race in (socio)linguistics. In turn, I highlight the benefits of "interdisciplinary thinking" (Charity Hudley, Mallinson & Bucholtz, 2020, p. e200) in sociolinguistic research involving race and centralize my experiences and insight as a Black linguist to draw attention to the importance of listeners' perception in Black criminalization processes. By doing this, the present study contributes to racial justice efforts in linguistics and importantly draws attention to listeners who racialize and criminalize "Black-sounding" speakers.

1.5 Dissertation Structure

This dissertation consists of eight chapters. Chapter 1 poses the topic, defines the research objectives, lays out the research question, and outlines the nature and scope of the present study. Chapter 2 historicizes this work by demonstrating how Blackness and criminality have been mutually co-constructed across time and space in America. In turn, I situate the present study within the enduring history of U.S. Black criminalization. Chapter 3 overviews and describes the method I used to collect speech samples from 100 speakers across the United States—the *Online Survey Speech Sampling* method. Then, Chapter 4 outlines my strategy to capture listeners' racialized categorizations of the speakers and speech samples–a process I call *voice racialization*. I also describe how I operationalized these categorizations to derive speakers' corresponding Black racialized speech scores (also *Black racialization scores*).

The second half of this dissertation turns to the empirical analyses of the speech samples. Chapter 5 is the first chapter analyzing speakers' production of speech, using statistical methods to determine whether some AAVE features predict speakers' Black racialization scores. I tested speakers' use of six AAVE phonological features and found a positive correlation between use of these features and speakers' Black racialization scores, but it was not statistically significant. Considering that this may have been because AAVE primarily represents the speech of a subset of Black speakers (King, 2020; Smitherman, 1988), in Chapter 6, I tested four additional AAL features. Yet, again, the speakers' Black racialization scores were not predicted by the extent to which they used these features. Taken together, these results suggest that listeners' racialization of speakers may be based more on their own listening practices (i.e., perception) rather than speakers' speech (i.e., production).

Next, Chapter 7 explains the memorization experiment I designed and deployed to investigate whether listeners associate Black racialized speech with criminality. The findings suggest that, even when listeners are given no visual cues to determine speakers' racial identities and do not otherwise know speakers' race, they may associate Black racialized speech with criminality. However, this association was only present among the listeners who heard Black speakers—although, as mentioned, the listeners were not told the racial identities of the speakers they heard. Finally, in Chapter 8, I review the previous chapters' findings, discuss the implications of the work, outline limitations, provide suggestions for future research and action, and present the conclusion.

2.0 Painting the Historical Landscape: History of Black Criminalization in America

2.1 Introduction

Taking into consideration the long and enduring history of Black criminalization in the United States, there are many opportunities for sociolinguists to investigate the ways in which linguistic practices contribute to Black criminalization in all facets of society. However, given the significance of Rickford & King's (2016) seminal work on Rachel Jeantel's testimony, much of the subsequent work on Blackness, language, and criminalization has focused on language in courtroom settings (especially defendant and witness testimonies) (e.g., Jones et al., 2019; Kurinec & Weaver III, 2019). In addition, Baugh's influential work on *linguistic profiling*—or the use of auditory cues to identify speakers' racial identities—is often contextualized to the courtroom, wherein linguistic profiling is believed to negatively impact Black defendants' outcomes when used as evidence in court (Baugh, 2000; 2005). While the present study builds upon the findings and theoretical contributions from earlier sociolinguists' research, I importantly situate it within the broader historical context of Black criminalization in America, which has taken place both within and outside of the courtroom (Davis, 2003; Hinton & Cook, 2021; Smiley & Fakunle, 2016). Accordingly, I aim to investigate the role of language in perpetuating Black criminalization in U.S. society more generally.

In this literature review, I summarize the history of Black criminalization in the United States from colonial times to the present. In doing so, I demonstrate the ways in which Blackness and criminality have been mutually co-constructed across time (Hinton & Cook, 2021). After, I give an overview of sociolinguistic research on Blackness, language, and criminality to show how

linguistic practices have also contributed to Black criminalization via U.S. courts (Corley [Branson], 2014; Jones et al., 2019; Kurinec & Weaver III, 2019; Rickford & King, 2016). In the discussion that follows, I explain how the historical contextualization of Black criminalization can benefit sociolinguistic research on the role of language in racialization and criminalization processes. Specifically, I highlight the need to supplement the current body of work on Black criminalization in courts with research that focuses on Black criminalization in other contexts and U.S. society more broadly. In doing so, I aim to expand sociolinguists' understanding of language's role in criminalizing Black Americans.

2.2 Black Criminalization from Colonial Times to the Present

From colonial times to the present, U.S. society has defined "criminality" in ways that are anti-Black and "Blackness" in ways that are criminalized. Thus, the mutual co-construction of Blackness and criminality has shaped what it means to be Black in America for the past 400 years (Hinton & Cook, 2021). The following comprehensive review of the United States' history of Black criminalization demonstrates that the practice is deeply rooted within American society and continues to impact Black Americans' experiences and social outcomes today.

2.2.1 Black Criminalization: Colonial Times through Emancipation

Historians have shown that the institution of slavery in the U.S. South is inextricably linked to the present-day carceral system in the United States (Boyles, 2015; Schrader, 2019). For example, jails and prisons first emerged in the south to house enslaved Africans prior to auction—

a practice that served to bolster the American institution of slavery while establishing an early link between Blackness and criminality (Henderson, 2016). In addition, jailers were sometimes called upon to inflict brutal punishment and other forms of torture on enslaved Africans and their descendants, thus creating an association between criminal legal actors and enslaved Africans even while few were formally incarcerated for breaking the law (Henderson, 2016). Slave patrols—a precursor to today's police force—helped to unite poor and wealthy southern White Americans who worked together to maintain the racial order by monitoring enslaved Black Americans' whereabouts, preventing uprisings, and inhibiting resistance efforts (Schwarz, 1998; Wagner, 2009). The passing of the Fugitive Slave Act in 1850, however, normalized and expanded these anti-Black policing practices across the Mason-Dixon line (Hinton & Cook, 2021; Walker, 1998). Thus, in the colonial south (and later in the north), the association of Blackness with criminality was established and reinforced through the institution of slavery, even while few enslaved southern Blacks were formally imprisoned for violating the law (Hinton & Cook, 2021).

Colonial times in the north marked the beginning of the modern penitentiary as well as the racial disparities that characterize it today (Hinton & Cook, 2021; McLennan, 2008; Rothman, 2017). For example, at Newgate Prison—New York's first state prison—25 percent of prisoners confined there from 1797 to 1828 were Black, despite the fact that Black New Yorkers were only 12 percent of the state's population. Similarly, in Philadelphia in 1816, 43 percent of incarcerated people identified as "Negro" or "Mulatto", even though these groups made up only 15 percent of Philadelphia's population (Hinton & Cook, 2021; Mishler, 2016; Nash, 1988). Taken together, during colonial times in the United States, the association between Blackness and criminality was established and reinforced predominately through anti-Black carceral practices in the north and the institution of slavery in the south.

2.2.2 Black Criminalization: Reconstruction and Jim Crow

While few enslaved southern Black Americans were incarcerated for violating the law, many were formally tried, convicted, and incarcerated via the criminal legal system during the period from the end of the Civil War to the passing of the Civil Rights Act—the Reconstruction and Jim Crow eras (Gross, 2006; Hicks, 2010: Taylor, 1938). This was primarily due to the establishment of Black Codes, or anti-Black laws put into place after Emancipation. While the Thirteenth Amendment extended a number of de facto rights to Black Americans, the Black Codes stripped these rights away (Gross, 2006; Hicks, 2010; Hinton & Cook, 2021; Taylor, 1938). They prohibited newly freed Black Americans from voting, owning arms, testifying in court, and importantly, "selling crops without permission from a White person" (Hinton & Cook, 2021, p. 268). Simply put, during the Reconstruction era, Black Americans were required by law to be employed by Whites. The codes, then, essentially forced recently freed Black Americans back into an exploitative slave-like labor system by allowing former enslavers to continue controlling Black labor while building their own wealth and growing influence in the south (Blackmon, 2009; Oshinksy, 1997). This, in turn, inhibited Black Americans' own efforts to create sustainable livelihoods and prompted many to flee southern oppression and seek gainful employment in northern and midwestern cities—a massive exodus of six million Black Americans called *The First Great Migration* (Derenoncourt, 2022).

However, many Black Americans who remained in the south and would not work for Whites were incarcerated and exploited anyway through "slavery by another name" (Blackmon, 2008), or convict leasing (Hinton & Cook, 2021). Rather than serving out their sentences in prisons, southern Black Americans who were convicted of violating Black Codes and frivolous vagrancy laws were hired out to do hard, undesirable labor like picking cotton, working in coal

mines and sawmills, and building railroads (Hinton & Cook, 2021; Oshinsky, 1997). Many Black adults subjected to convict leasing did not survive a ten-year sentence (Oshinsky, 1997). Even Black children were not exempt from the inhumane practice of convict leasing. Black boys at Dozier School for Boys in Marianna, Florida, for example, were regularly leased out to work for local plantation owners. Nearly 100 boys died at Dozier before the reform school was finally shut down in 2011 (Robinson, 2022).

Post-slavery efforts to control Black labor also contributed to a rise in White supremacist extralegal forms of justice and mob lynching during the Reconstruction and Jim Crow eras (Dray, 2002; Wells, 1895). Even before the turn of the 20th century, thousands of Black Americans were extrajudicially lynched for *alleged* murder, rape, well poisoning, race prejudice, and even "no offense" (Wells, 1895). Though Black Codes and convict leasing were legal practices in the south, mob lynching was not. Yet, law enforcement officers often corroborated with mobs to execute Black citizens accused of crimes, and because of this, the perpetrators were rarely held accountable (Dray, 2002). And while the horrors of lynching have been meticulously documented since the late 1800s (Wells, 1895), the first bill to make lynching a federal hate crime was not signed into law until March of 2022 after having been shot down continuously for over 100 years (18 U.S.C. § 249, 2022).

2.2.3 Black Criminalization: Civil Rights Era and Wars on Crime and Drugs

Jim Crow laws were outlawed with the passing of the Civil Rights Act in 1964 (Packard, 2003). However, in much the same way that Black Codes and convict leasing worked to preserve slavery after it was abolished, the wars on crime and drugs contributed to the emergence of a new Jim Crow that perpetuated the criminalization of Black Americans in U.S. society (Alexander,

2020). For example, just nine months after the passing of the Civil Rights Act, President Lyndon Johnson waged a war on crime across America. His war on crime aimed to "improv[e] police-community relations and crack…down on neighborhood disorder" with proactive policing reform—specifically, through the development of militarized surveillance strategies to analyze, monitor, and predict crime (Hinton & Cook, 2021, p. 274). Yet, these techniques encroached on Black Americans' legal rights and fortified Americans' association of Civil Rights' advocacy with violence and crime.

Similarly, the "War on Drugs," beginning during President Richard Nixon's term in the 1970s and continuing through President Ronald Reagan's term to the end of the 1980s, exacerbated racial disparities in the U.S. criminal legal system that were already growing in the decades before (Provine, 2011). The emergence of crack cocaine as an affordable alternative to cocaine powder inspired Nixon's tough-on-crime policies, but disproportionately targeted residents in poor Black neighborhoods (Sirin, 2011). Thus, the increase in convictions issued to Black Americans during this period worked to reinforce racial inequality and the racial hierarchy. Echoing the Jim Crow era, the wide-spread criminalization of Black Americans mongered fear, reduced rights (e.g., voting and jury eligibility, access to public benefits, freedom from housing and employment discrimination), and curtailed economic opportunities. In short, Black Americans remained second-class citizens despite now having full legal rights (Alexander, 2020).

Moreover, during this period, criminological research was used to justify racially biased crime policies and policing strategies. For example, Kelling and Wilson's (1982) Broken Windows theory, was instrumental in formulating Reagan's proactive crime prevention strategies. The theory put forth that "fear of violent crime and societal dysfunction directly correlated with the prevalence of broken windows, or minor criminal offenses, if left unchallenged in neglected

communities" (Hinton & Cook, 2021, p. 276). However, broken windows policing ultimately led to increased police presence and disparate arrests in urban Black neighborhoods without necessarily improving crime rates (Harcourt, 2005). Due to the clear connection between the wars on crime and drugs, the striking racial disparities in America's criminal legal system, and the impact of a criminal record, this era from the 1960s-1990s is often associated with mass incarceration of Black Americans (Hinton & Cook, 2021, Pager, 2003). However, striking racial disparities and the impact of a criminal record continue to be problems in American society today (Alexander, 2020; Pager, 2003).

2.2.4 Black Criminalization: 1990s – Black Lives Matter Era

Persistent racial disparities in profiling, arrests, guilty verdicts, and incarceration rates evidence the deeply rooted association between Blackness and criminality in the U.S. criminal legal system (Alexander, 2020; Hinton & Cook, 2021; Johnson & Lee, 2013; Kovera, 2019; Pager, 2003; Sommers & Marotta, 2014; Yang, 2015), Yet, *#LivingWhileBlack* incidents from the 1990s to the present also show that Blackness is associated with criminality beyond the criminal legal system proper and all throughout society (Thurston, 2019). *#LivingWhileBlack* is a widely used hashtag popularized by Black American writer, activist, and comedian Baratunde Thurston. The hashtag describes encounters in which a person (often a White person) feels uncomfortable or threatened by a Black person doing a harmless, everyday activity, and responds by calling the police or themselves using other excessive or deadly force (Thurston, 2019). In 1991, for example, Latasha Harlins (15 years old) was shot from behind and killed by a shopkeeper while purchasing orange justice. In 2012, Trayvon Martin (17 years old) was shot and killed by a neighborhood watchman while walking in his father's fiancée's neighborhood. And in 2020, Ahmaud Arbery

(25 years old) was shot and killed by three White men while he was jogging. In the cases of the two teens, Harlins' killer received 400 hours of community service, a $500 fine, and probation and Martin's killer was found *not guilty*. These incidents eerily resemble the extrajudicial lynching of Black Americans during the Reconstruction and Jim Crow eras and the outcomes of these cases reflect the society's blatant disregard for Black lives.

While deadly force is not always used in *#LivingWhileBlack* encounters, it is often summoned via the police. Black adults and children have had the police called on them for being "agitated" while walking to work, using the wrong type of barbeque grill, playing golf too slowly, packing their cars, waiting for an Uber, canvassing in their own neighborhood, attempting to cash a check, using a coupon, going back and forth to the bathroom, operating a community garden, using a swimming pool, reading C.S. Lewis, donating food to people who are homeless, sitting in Starbucks, and resting their feet on a chair in a classroom (Thurston, 2019). One 53-year-old White woman called the police on a nine-year-old Black boy for sexual assault when his backpack accidentally touched her as he walked by (Thurston, 2019). These incidents in U.S. society, along with continued racial disparities in the American criminal legal system, show vividly that Blackness and criminality are mutually co-constructed to reinforce Black Americans' low positioning on the racial hierarchy and promulgate White supremacy (Hinton & Cook, 2021).

Taken together, for over 400 years, Blackness has been associated with criminality in the United States. During colonial times, Blackness was entrenched with criminality, even as Blackness itself emerged as a social construct. In addition, the association of Blackness with criminality was reinforced by the normalization of hyper-surveilling Black Americans (before and) after the passing of the Fugitive Slave Act in 1850 as well as the devaluation of Black bodies through convict leasing and Black Codes after Emancipation. Wars on crime and drugs perpetuated

the association further from the 1960s-1990s. Moreover, the fact that, today, White Americans call the police on Black Americans and shoot and kill them for everyday behaviors (e.g., buying orange juice, walking, and running) shows that Blackness is associated with criminality in the United States. This deeply rooted association of Blackness with criminality continues to reinforce racial hierarchies and reify the inequalities prevalent at every step in the criminal legal process and U.S. society more broadly.

2.3 Black Criminalization Research in Sociolinguistics

Given the long and enduring association of Blackness with criminality, since the 1990s, sociolinguists have investigated the role of language in perpetuating and reinforcing Black criminalization. However, despite the large, coordinated research efforts to address Black children's language needs in U.S. schools from the 1960s to the present (Baker-Bell, 2020a; Baker-Bell, 2020b; Baugh, 1979; Fasold, 1972; King, 2020; Labov, 1966; Rickford, 1999; Rickford, Sweetland & Rickford, 2004; Rickford, 2005; Wolfram, 1969), outside of a few dedicated sociolinguists, the issue of Black criminalization has not, until recently, received the same attention or resources in sociolinguists. In fact, in 2010, leading forensic linguist Diana Eades pointed out that there is, "almost no linguistic research which examines African Americans' interactions in the legal process" (p. 89). Her caveat, "*almost* no…research," likely refers to John Baugh's (1996, 2000, 2005) work on linguistic profiling, which until the mid-2010s was the primary research focusing on Black Americans' language in legal settings—specifically, in courts.

Linguistic profiling refers to the use of auditory cues to identify a person's racial identity and is the auditory counterpart to "racial profiling" (Baugh, 1996; 2000; 2005, p. 155). Baugh's

work on linguistic profiling is in large part based on his earlier research showing that listeners are remarkably good at identifying speakers' racial identities based on voice alone (Purnell, Idsardi & Baugh, 1999). He postulated that listeners do this by using auditory cues to discern language varieties that are linked to racial identities (Purnell, Idsardi & Baugh, 1999). In his work on linguistic profiling, he highlights that the practice may impact Black speakers' legal outcomes when auditory cues are used as evidence in court (Baugh, 2005). For example, in *Clifford v. Kentucky* (1999), the Kentucky Supreme Court ruled that linguistic profiling evidence submitted by a White police officer who identified a voice as a "Black male" (i.e., the defendant) was *not* unconstitutional. Therefore, the court upheld the defendant's prior guilty verdict which was largely based on the linguistic profiling evidence (Baugh, 2005, p. 156).

While jurisdictions across the United States have ruled differently on the constitutionality of admitting linguistic profiling evidence in court, Baugh's work importantly highlights that race is not only seen but *heard* and, like racial profiling, has important consequences for racialized speakers in criminal legal settings (2000; 2005; 2018). Baugh remained the primary voice on Black language in legal settings until controversy surrounding a testimony in a high-profile 2013 criminal case, *State of Florida v. George Zimmerman*, ignited new energy around the issue and led other sociolinguists—including prominent Black linguist and AAVE advocate, John Rickford, Sr., and burgeoning AAVE dialectologist, Sharese King—to investigate it (Corley [Branson], 2014; Jones et al., 2019; Kurinec & Weaver III, 2019; Rickford & King, 2016).

The testimony of Rachel Jeantel, a 19-year-old Haitian American woman, friend of Trayvon Martin, and star witness for the prosecution in the *State of Florida v. George Zimmerman* trial, is at the center of sociolinguistic research on Black language in legal settings from the past 10 years (Corley [Branson], 2014; Jones et al., 2019; Kurinec & Weaver III, 2019; Rickford &

King, 2016). Jeantel was on the phone with Martin moments before he was shot and killed by Zimmerman, and she recounted her final interactions with Martin for nearly six hours on the stand, using AAVE to give her statements. Yet, her crucial testimony in this important case was dismissed by jurors as "incomprehensible" and "not credible" (Rickford & King, 2016, p. 948). I (2014) along with prominent Black sociolinguists, Rickford and King (2016), psycholinguists Kurinec and Weaver III (2019), and sociolinguists Jones et al., (2019) have postulated that the jurors' ousting of her testimony may have contributed to Zimmerman's *not guilty* verdict. Qualitative (Rickford & King, 2016) and experimental (Corley [Branson], 2014; Jones et al., 2019; Kurinec & Weaver III, 2019) research methods across multiple linguistic subdisciplines (e.g., sociolinguistics, psycholinguistics) were used to draw this conclusion.

In 2016, Rickford and King (2016) conducted a qualitative assessment of Jeantel's testimony. They examined Jeantel's grammar (morphosyntax), phonetics and phonology, and lexicon and concluded that she spoke a systematic linguistic variety that is primarily based on AAVE but has some influence from Haitian Kweyol or Caribbean English (Rickford & King, 2016). Based on their analysis, they attributed the jurors' assessments of Jeantel's credibility and intelligibility to "...(i) dialect difference and unfamiliarity; (ii) Jeantel's underbite and voice quality; (iii) and attitudes including dialect bias and institutionalized racism/prejudgment" (Rickford & King, 2016, p. 972). Thus, the authors drew a link between the jurors' incomprehensibility and negative views of AAVE and their assessments of the credibility of Jeantel's testimony.

In 2014, I conducted a matched-guise study using 24 Cornell University undergraduate students as mock jurors. Half of the participants heard two minutes of Jeantel's original AAVE testimony and the other half heard a two-minute sdAE version I created. I tested whether the mock

jurors' assessments of Jeantel and her testimony differed across the language varieties and found that they understood less of the substantive content of the AAVE testimony compared to the one in sdAE, and they evaluated the witness as less informative and less reliable in AAVE compared to sdAE (Corley [Branson], 2014). Rickford and King's (2016) study inspired two follow-up quantitative experiments by Kurinec and Weaver III (2019) and Jones et al. (2019). Kurinec and Weaver III (2019) employed a similar matched-guise technique to my 2014 study. However, they created four guises (General American English [GAE] defense witness, AAVE defense witness, General American English defendant, AAVE defendant) each in three different evidentiary contexts (ambiguous, pro-prosecution, pro-defense) (Kurinec & Weaver III, 2019). The AAVE speaking witness in their study was evaluated as "less professional" and "less educated" in all contexts and was more likely to be found guilty compared to the GAE speaking defendant, except when the case materials favored the prosecution (Kurinec & Weaver III, 2019, p. 803). Jones et al.'s (2019) study tested Philadelphia court reporters' abilities to accurately transcribe and paraphrase AAVE testimony. The court reporters fell well below the required 95 percent accuracy rate and their mistranscriptions reflected implicit negative views of AAVE and Black speakers. The authors also noted that court reporters "immediately recognized [AAE] voices as 'black' and associated these voices with criminality, deviance, and untrustworthiness" (Jones et al., 2019, p. e247). Taken together, this body of qualitative and experimental work importantly highlights that listeners' biases towards AAVE may contribute to unfair treatment of Black speakers and increase their likelihood of being perceived as 'guilty' or criminalized in court settings.

2.4 Discussion

The research on Blackness, language, and criminality summarized above centralizes language and linguistic racism in courts. And while the perpetuation of Black criminalization via linguistic processes/practices in the courtroom (i.e., defendant and witness testimonies) necessitates sociolinguistic investigation, only five to ten percent of criminal cases go to trial. The majority are resolved by plea bargaining and, thus, defendants' guilt and witnesses' believability (i.e., criminality) is not assessed by a jury (Herman, 2012, Neubauer & Fradella, 2018). Moreover, the historical overview presented earlier explains and demonstrates how Black criminalization has taken place, and still occurs, in other areas of the criminal legal system and all throughout U.S. society. This highlights the need to examine the relationship between Blackness, language, and criminality in other non-court contexts. Preliminary sociolinguistic inquiry on this relationship also emphasizes this (e.g., Holliday, Burdin & Tyler, 2015).

In a 2015 blog post, for example, linguists Holliday, Burdin, and Tyler posited that the way in which an officer's perception and assessment of a Black woman's voice as "combative" may have influenced his criminalization of her as well as his use of excessive force in the woman's arrest, which arguably led to her death. This blog post helps to illuminate how linguistic processes may contribute to Black criminalization in areas outside of the courtroom and systematic sociolinguistic investigation is needed to determine this. By expanding beyond courtroom contexts, sociolinguists have an important opportunity to assess the ways in which Blackness and criminality are mutually co-constructed through profiling, policing, incarceration, post-incarceration disparities, and in other institutions, organizations, neighborhoods, public spaces, and U.S. society in general (Alexander, 2020; Hinton & Cook, 2021; Johnson & Lee, 2013; Kovera, 2019; Pager, 2003; Sommers & Marotta, 2014; Thurston, 2019; Yang, 2015). To more

fully understand how linguistic practices contribute to Black criminalization, sociolinguistic research should consider linguistic practices in these other contexts.

The present study aims to do this by testing whether U.S. listeners associate Black-sounding speech (i.e., Black racialized speech) with five different representations of criminality. I use a psycholinguistic memorization experiment to test whether listeners associate Black racialized speech with crime relevant images (e.g., gun, handcuffs)[2], which represent different manifestations of criminality within and outside of the courtroom. Therefore, by conducting the experiment this way, I attempt to situate the present study within the broader historical context of Black criminalization in America–which occurs in many different contexts. In turn, the findings apply to multiple social contexts in which Black Americans are criminalized, and the work can be used to address Black criminalization in institutions like hospitals, schools, and the housing industry as well as in everyday interactions like *#LivingWhileBlack* incidents–in addition to courts.

2.5 Conclusion

Criminalization has characterized what it means to be Black in America since colonial times. Therefore, today, sociolinguists have an important opportunity to examine the role of language in perpetuating Black criminality in the United States. By taking into account the long history of Black criminalization and building upon the current body of research on Blackness, language, and criminality, the present study investigates how linguistic practices contribute to Black criminalization in other areas of the criminal legal system and society more broadly. As

[2] This research experiment and its results are explained in detail in Chapter 7.

such, the research findings can apply to a range of situations and contexts where Black Americans are criminalized. In the following chapters, I outline the processes I used to collect speech samples, categorize Black racialized speech, and conduct the psycholinguistic memorization experiment to test my overarching research question, *Do U.S. listeners associate Black racialized speech with criminality?*

3.0 Data Collection: Increasing Black Representation in Spoken Language Corpora

3.1 Introduction

Over the years, linguists have utilized many novel speech sampling strategies to acquire the audio recordings they use in their analyses (Hernández-Campoy, 2014). In general, these strategies fall into two categories: supervised and unsupervised. With supervised methods, speakers meet with researchers in-person (or virtually) to have their speech recorded professionally (Poorjam & Jensen, 2019). With unsupervised methods, however, speakers use web-based interfaces, dedicated phone-lines, smart device applications, wear a recording device, or employ another strategy to independently record and submit speech samples remotely, without direct assistance from the researcher (Poorjam & Jensen, 2019). Because unsupervised methods do not require one-on-one interfacing between speakers and researchers, they are simpler, faster, and better suited for building spoken language corpora to use in large-scale quantitative research compared to supervised methods (Poorjam & Jensen, 2019; Lane et al., 2010). Accordingly, over the past 30 years, linguists have innovated and implemented unsupervised data collection methods to build large corpora of spoken language for a wide range of linguistic purposes (Gruenstein, 2009; Godfrey, Holliman & McDaniel, 1992; Nagrani et al., 2017).

Recently, for example, researchers have used these corpora to develop automated speech recognition (ASR) models (Lane et al., 2010), which are the basis for systems that automatically detect and transcribe speech in real-time (Rawat, Gupta & Kumar, 2014). However, some sociolinguists have pointed out that these systems perform much worse for Black compared to White speakers. They have also illuminated that this disparity reflects the fact that the corpora

from which ASR models are developed have far fewer samples of speech from Black compared to White speakers and the Black speech samples that are included are not very diverse (Koenecke et al., 2020; Martin, 2022; Mengesha et al., 2021). This limits the use of these corpora not only for developing ASR models, but also other large-scale quantitative research involving Black speakers like, studying variation among Black speakers across the United States, describing cross-regional Black speech practices, and investigating listeners' perception of Black speakers' voices and speech.

3.2 Increasing Black Representation in Spoken Language Corpora

Since 2018, sociolinguists have begun to address the lack of Black representation in spoken language corpora by deploying coordinated efforts to build a database of spoken African American Language (AAL)—the Corpus of Regional African American Language (CORAAL) (Kendall & Farrington, 2021). Yet, CORAAL's recordings are acquired via supervised methods (i.e., sociolinguistic interviews) and in some communities, have taken years to collect (Kendall & Farrington, 2021). Thus, there is a need for unsupervised/remote data collection methods to build large corpora of spoken language from Black speakers more efficiently, not only for improving ASR models, but for representative large-scale quantitative research involving Black speakers in general. Such a method would address both limitations of existing unsupervised corpora, which currently have too few Black speakers and not enough diversity among the speech samples from Black speakers. To fill this gap, I deployed an unsupervised remote speech sampling method, which I call *Online Survey Speech Sampling*, to build a corpus of spoken language from diverse

Black speakers. The recordings from this corpus were used to test listeners' perception of speakers' racial identities and whether they associated the speakers' voices with criminality.

Briefly, the approach integrates three online platforms to recruit a representative sample of Black speakers, collect recordings through an online survey, and automatically transcribe the recordings completely remotely, without researcher supervision.[3] In what follows, I describe how I implemented the Online Survey Speech Sampling method to build a corpus of spoken language from Black American speakers—as well as from Non-Black U.S. speakers for comparison. I then assess the method's effectiveness by evaluating the quality of the speech samples and determining whether the method captured speech from representative cross-sections of Black and Non-Black U.S. speakers, which would presumably increase the diversity of speech samples represented in the corpus. Finally, I discuss the results, their implications, and conclude.

3.3 The Online Survey Speech Sampling Method

The Online Survey Speech Sampling method integrates three internet-based technological platforms to collect speech samples from a representative cross-section of Black and Non-Black U.S.-based speakers completely online and without researcher supervision. Using an internet-based approach does limit the speech sampling to U.S. speakers with internet; however, 93 percent of Americans have internet access (Pew Research Center, 2021). In addition, similar approaches have integrated online platforms to conduct unsupervised linguistic experiments (e.g., Garcia,

[3] Given ASR's limitations regarding automatic transcription of Black speech, I needed to make hand-corrections to the transcripts prior to using them in later analyses. (This process is described in Chapter 5 and Chapter 6.) Despite problems with the automated transcripts, making corrections by hand was much more efficient than transcribing the audio files from scratch.

Roeser & Kidd, 2021) and to collect speech samples via *supervised* methods (e.g., Leemann et al., 2020). But the Online Survey Speech Sampling method differs from these in that it integrates the online platforms to collect speech samples using *unsupervised* methods and is, therefore, faster and more efficient. The three integrated technological platforms are *Qualtrics*, *Phonic*, and *Prolific*, and each plays a critical role in making the speech sampling approach feasible and efficient. Below, I describe each platform and its role in the online speech sampling process.

3.3.1 Qualtrics

Qualtrics is an online survey building site (Qualtrics, Provo, UT). The website offers survey templates and tools to build completely customizable online questionnaires. As an internet-based platform, Qualtrics surveys can be developed and deployed to participants completely online, with no interface between researchers and participants—thus increasing the breadth of speakers who can participate as well as decreasing the time it takes to collect recordings and create spoken language corpora.

3.3.2 Phonic

Phonic (www.phonic.ai) is an online platform for recording voices and creating videos. The site allows researchers to embed voice or video recording widgets, or *plug-ins*, to Qualtrics or other online surveys. With Phonic, research participants can record their voices into surveys directly, without the extra steps of recording an audio file, saving it, and uploading it to the survey. Speakers can listen to their recordings and re-record as many times as they please. Only the last version, however, is saved and uploaded to the researcher's account. Once the audio files are

uploaded, Phonic provides an automated verbatim transcript of the recording. So, researchers have raw audio files as well as transcribed text for each submitted recording. Phonic is also affordable, costing academic researchers just 65 dollars per month (and 129 dollars per month for non-academic researchers) to elicit an unlimited number of recordings.

3.3.3 Prolific

Prolific (www.prolific.co) is an online participant recruitment site. It hosts thousands of research participants who receive payment for participating in research studies. Researchers can recruit participants from specific demographic backgrounds by using the site's participant prescreening filters. By selecting "Black/African American," the Qualtrics survey is advertised only to Prolific participants who self-identify as 'Black'[4] and by selecting all other racial identities except for 'Black,' the Qualtrics survey is advertised only to Prolific participants who do not self-identify as Black.

Taken together, the Online Survey Speech Sampling method works by using Prolific to recruit speakers from a specified population who complete a Qualtrics survey and record their speech using embedded Phonic widgets. I have conveniently outlined the process I used to integrate Qualtrics, Phonic, and Prolific, including all HTML codes, and this information is available in Appendix A.

[4] As of June 8, 2023, Prolific had 2,879 Black/African American-identifying participants on their site.

3.4 Survey Development and Speech Sample Collection Processes

I used the Online Survey Speech Sampling method to collect 500 speech samples from 50 Black and 50 Non-Black U.S. based speakers. To implement the Online Survey Speech Sampling method, I embedded five Phonic voice recording widgets into a Qualtrics survey and distributed the survey to 100 Prolific participants (henceforth *speakers*). To elicit naturalistic, rather than scripted speech, speakers watched four short video clips of a person making a peanut butter and jelly sandwich and narrated what they saw happening in each video clip.[5] The corresponding audio files were labeled "PBJ1", "PBJ2", "PBJ3", and "PBJ4", to represent each step in the peanut butter and jelly sandwich-making process. Speakers also recorded themselves counting from 1-20 and these files were labeled "1-20". By employing these elicitation strategies, I collected 500 samples of naturalistic speech without ever interfacing with speakers directly, allowing me to efficiently build a corpus of spoken language recordings from Black and Non-Black speakers, and in later chapters, I explain how I used the samples from this corpus in multiple sociolinguistic analyses and experiments.

To assess whether the speech samples met quality standards, I conducted a listening assessment using the criteria in the table below. I chose to analyze only 200 of the 500 speech samples (100 PBJ4 and 100 1-20), as I focused the analysis and discussion in the dissertation on this subset of the data. I also ran a series of *t*-tests on these 200 speech samples to assess whether my samples of speakers were representative of their counterparts in the general population across age, sex, education level, income level, and U.S. division of residence (proportions were obtained from The American Community Survey, 2021).

[5] The videos I used are publicly available at https://www.wikihow.com/Make-a-Peanut-Butter-and-Jelly-Sandwich.

Table 1 Audio Quality Assessment Criteria

Score	Rating	Background Noise	Audio Popping or Buzzing	Speaker Clarity
4	Excellent	None	None	Clear
3	Good	A little—only noticeably with careful listening	A little—only noticeable with careful listening	Clear
2	Satisfactory	Some—noticeable without careful listening	Some—noticeable without careful listening	Clear
1	Poor	Substantial	Substantial	Clear
0	Unusable	Substantial	Substantial	Not Clear

3.5 Results

One clear advantage of using this method is the speed and efficiency at which data can be collected. In this case, I distributed the Qualtrics survey at 7:05am on May 5, 2022 and had all 500 recordings just 8 days later by May 13, 2022, at 5:48am. I assessed the quality of the speech sample recordings by listening to each of the 200 speech samples slowly and carefully and rating them using the criteria outlined above. The results of my assessment are in the table below.

Table 2 Audio Quality Assessment Results

Score	**Rating**	**Proportion of Clips with Corresponding Rating**
4	Excellent	0.77
3	Good	0.14
2	Satisfactory	0.06
1	Poor	0.03
0	Unusable	0.00

I used *R* (R Core Team, 2020) to run 42 *t*-tests to determine whether the samples of Black and Non-Black speakers were nationally representative across age, sex, education level, income

level, and U.S. division of residence. The results of these tests are provided in the table below, and speaker characteristics with an asterisk (*) indicate statistical significance. Statistical significance indicates that the sample of speakers is *not* nationally representative along that particular demographic.

Table 3 T-Test Results Comparing Speakers' Characteristics to National Proportions

	Black Speakers	**Non-Black Speakers**
Speaker Characteristics		
Age	38.02*	39.68
Sex		
Female	0.48	0.46
Male	0.52	0.54
Education		
Less Than High School	0.02*	0.02*
High School	0.20*	0.08*
Associate's, Trade School, Some College	0.34	0.28
Bachelor's Degree	0.32*	0.40*
Graduate Degree	0.12	0.22
Income		
Less Than $25,000	0.18	0.16
$25,000 - $50,000	0.36	0.24
$50,000 - $100,000	0.38	0.36
Over $100,000	0.04*	0.20*
U.S. Division		
New England	0.06	0.04
Middle Atlantic	0.06*	0.24
Eastern Midwest	0.10	0.18
Western Midwest	0.04	0.12
Eastern South	0.38	0.24
Deep South	0.10	0.00
Western South	0.14	0.06
Mountain	0.04	0.06
Pacific	0.08	0.06*
Number of Speakers	**50**	**50**

*Denotes that the coefficient is statistically significantly different from zero with a p-value <0.05

3.6 Discussion

Combined, the results of the analyses show that the Online Survey Speech Sampling method was moderately successful at producing quality speech samples from a representative cross-section of Black and Non-Black speakers. Specifically, the results of the listening assessment show that most of the speech samples (91 percent) are good or better quality. Only six of the 200 audio clips are poor; yet because the speakers in these recordings could still be heard clearly, the speech samples were deemed suitable for use in the later sociolinguistic experiments and analyses. In addition, the results of the *t*-tests showed that the samples of speakers are mostly representative of their corresponding national populations across sex, income, and U.S. division of residence. The sample of Non-Black speakers is also representative in age; however, Black speakers in the sample are older than Black Americans in the population more generally. Finally, both samples are more educated than their counterparts in the general population. These findings are likely explained by the fact that Prolific does not permit anyone under the age of 18 to participate in research studies on their site. Therefore, the samples represent Black and Non-Black adults only—who are older and thus have higher education levels than their corresponding national groups overall.

While the method boasts many advantages, the disadvantages are that it is not participatory, and the automated Phonic transcripts contain errors. First, in the past decade, sociolinguists have advocated for the use of participatory research methods to better inform research involving Black speakers and increase the benefits of the research to these speakers and their communities (Wolfram, 2023). However, the Online Survey Speech Sampling method does not involve input from speakers and is, therefore, limited in this regard. Second, the automated transcriptions produced by Phonic contained errors that needed to be corrected by hand before use in the later

studies. This limits the effectiveness of the method for larger samples; however, it was suitable for the 200 samples assessed in this chapter. Taken together, though, the method was simple, effective, and practical and allowed me to collect 250 speech samples each from generally representative cross-sections of Black and Non-Black speakers in about eight days. Given the lack of representation of Black speakers in spoken language corpora, the Online Survey Speech Sampling method could be easily deployed to increase Black speakers' representation in these corpora and, accordingly, in large-scale quantitative linguistic research.

3.7 Conclusion

I implemented the Online Survey Speech Sampling method to develop a speech corpus that included samples from diverse Black speakers. I used the method to collect 500 speech samples from Black and Non-Black speakers to determine how well it performed in terms of obtaining quality speech samples from representative cross-sections of speakers. While I obtained high quality speech samples using the method, the speakers were not entirely representative of their corresponding national populations. Yet, the speakers were representative across sex, income, and U.S. division of residence, and thus, represent a wide range of speech practices along these dimensions. Therefore, the method succeeded with regards to the aims for which it was deployed and, importantly, is capable of increasing the representation of Black speakers and their diverse speech practices in spoken language corpora.

4.0 Hearing Black: Operationalizing Listeners' Perception of Black-Sounding Speech

4.1 Introduction

When listeners perceive speech, they interpret what they hear and may also make judgments about speakers' social identities—for example, they may categorize speakers into racialized, gendered, or sexual identity groups. Research shows that this is the case even when listeners do not explicitly know, or have other external cues to decipher, this information (Baugh, 1996; Gaudio, 1994; Gelfer & Mikos, 2005; Levon, 2007; Purnell, Idsardi & Baugh, 1999; Thomas & Reaser, 2004). Sociolinguists explain this finding by asserting that listeners attend to speakers' vocal production features to draw conclusions about their identity groups (Purnell, Idsardi & Baugh, 1999; Levon, 2007; Gelfer & Mikos, 2005; Thomas & Reaser, 2004). In other words, sociolinguists contend that listeners categorize speakers into social groups based on their use (or non-use) of racialized, gendered, or sexualized (for example) language features.

However, emerging work in raciolinguistics and psycholinguistics has identified a disparity between speakers' vocal production of speech and what listeners 'hear' in cases where listeners have additional information about speakers' identities (Flores & Rosa, 2015; Hay & Drager, 2010; Inoue, 2006; Levon, 2014; Niedzielski, 1999; Weissler, 2021). In fact, the work has shown that when listeners have this additional information, the ways they categorize and assess speakers/speech does not seem to be based on speech alone. Put differently, other, non-speech production-based factors appear to influence how listeners hear, categorize, and assess speakers and their speech in situations where listeners have additional, non-linguistic information about the

speakers (Flores & Rosa, 2015; Hay & Drager, 2010; Inoue, 2006; Levon, 2014; Niedzielski, 1999; Weissler, 2021).

Raciolinguists—researchers who interrogate the mutual co-construction of language and race—and some scholars in similar critical linguistic subfields, have utilized literary and philosophical theories to build a compelling case that listeners' perception of speakers/speech is significantly influenced by external social factors that, in turn, shape how listeners hear, categorize, and assess speakers/speech (Alim et al., 2016 Flores & Rosa, 2015; Inoue, 2006). Additionally, psycholinguists have conducted experiments showing that visual cues and listeners' prejudgments of speakers predict listeners' perception of speech separate from speakers' objective vocal production of it (Hay & Drager, 2010; Levon, 2007; Niedzielski, 1999; Rubin, 1992; Weissler, 2021). Taken together, work in raciolinguistics (and related critical linguistic subfields) and psycholinguistics highlights that the ways in which listeners hear, categorize, and assess speakers/speech are not based on speakers' production of speech alone. In fact, listeners' own perceptual biases play an important, if not critical, role. Yet, what is the role of perception when listeners hear, categorize, and assess speakers/speech without additional information about speakers' identities? In this chapter, I begin to address this question by deriving and implementing a perception-based heuristic in order to start teasing apart speakers' production from listeners' perception in speaker/speech categorization processes—specifically, when listeners do not have additional information about speakers.

First, I outline the need for listeners' categorization measures to examine the role of listeners' perception in speaker/speech categorization processes. Then, I describe how I derived and implemented a perception-based heuristic to interrogate the role of listeners' perception in their categorization and assessment of disembodied speech. This measure combines multiple

listeners' categorizations of speakers/speech samples into socially constructed groups and thus represents the likelihood a listener will categorize a particular speaker into a particular social group. I used the measure in a test case on racial categorization by speech alone. To do this, I collected speech samples from 100 U.S. speakers, obtained multiple listeners' categorization of these speakers into racialized groups—a process I call *voice racialization*—and calculated the proportion of listeners who categorized each speaker as "sounding Black." In turn, each speaker was given a Black racialization score, which represents the likelihood the speaker is categorized as "sounding Black" by listeners. After calculating these scores, I conducted linear regression tests to evaluate whether speakers' self-identified race (on a Black-non-Black binary) predicted the likelihood they were racialized as "sounding Black" by listeners. Throughout the remainder of this chapter, I describe the method and results and discuss the implications of my findings. In particular, I outline how the speech samples and corresponding Black racialization scores are used in an experiment in Chapter 7 to evaluate whether "Black-sounding" speech is associated with criminality.

4.2 Measuring Speech Categorization

4.2.1 A Mounting Case for Perception

For decades, many sociolinguists have proposed that when listeners do not have additional external cues to decipher speakers' social identities, they draw from features in the speech they are hearing to ascertain this information (Baugh, 1996; Gelfer & Mikos, 2005; Levon, 2007; Purnell, Idsardi & Baugh, 1999; Thomas & Reaser, 2004). However, some scholars who have attempted

to isolate the speech features that linked speakers to their identities were unable to do so (e.g., Gaither et al., 2015; Levon, 2014; Weissler, 2021). In the case of racial categorization, for example, Gaither et al. (2015) examined morpho-syntactic, phonological, and phonetic features in "Black-" and "White-sounding" speech but did not find that differential use of these features predicted whether speakers "sounded Black" or "White" to listeners. While Gaither et al.'s study was based on listeners' assessments of speech alone, it appears that speakers' production of speech may not have been the only factor shaping how the listeners perceived and categorized the speakers.

Other studies have drawn similar conclusions about the ways in which non-production-based factors influence listeners' perception, categorization, and assessment of speakers/speech (e.g., Hay & Drager, 2010; Levon, 2014; Niedzielski, 1999; Rubin, 1992). These studies, however, included additional non-production-based variables (e.g., explicit information about speakers' identities) to test whether this information influenced listeners' perception. Over and again, researchers found that these non-production-based factors indeed shaped listeners' perception (e.g., Hay & Drager, 2010; Levon, 2014; Niedzielski, 1999; Rubin, 1992). Although the present study focuses on racial identification by speech alone, the conclusions drawn from the work using non-production-based factors nevertheless helps to illuminate how these factors may still shape listeners' perception, categorization, and assessment of speakers/speech even when listeners' do not have access to this information. This research incorporating non-production-based factors has generally taken two approaches to draw conclusions about listeners' perception—experimental and philosophical.

4.2.1.1 Experimental Approaches

First, some linguistic researchers have utilized sociophonetic experiments to try disentangling production features from listeners' perceptions (e.g., Gaither et al., 2015; Levon,

2007). For example, Levon (2007) digitally manipulated *s* sounds in a speech sample to test whether lispy-er patterns correlated with listeners' assessments that the speaker "sounded gay." He found that listeners' prejudgments of the speaker's sexual orientation dictated whether they thought he "sounded gay," regardless of the degree of lisping (Levon, 2007; 2014). Likewise, another study by Rubin (1992) used a psycholinguistic experiment to investigate undergraduates' assessments of racially minoritized teaching assistants. He found that undergraduate students rated an Asian teaching assistant as more accented and a poorer instructor than a White teaching assistant even though the speech stimuli were identical across both experiment conditions. Together, these psycholinguistic experiments and others like them have built a convincing empirical case highlighting the importance of perceptual biases in listeners' speech categorization processes separate from speakers' production of speech (e.g., Drager, 2010; Hay & Drager, 2010; Weissler, 2021).

4.2.1.2 Philosophical Approaches

Second, other scholars, like raciolinguists, have drawn from literary and philosophical observations to theorize why socially constructed categorization is driven by listeners' perception, not the vocal production of speech (Flores & Rosa, 2015; Inoue, 2006; Levon, 2014; Paris & Alim, 2014). In fact, they emphasize the importance of listeners' *racializing* listening practices over *racialized* speakers' speech practices and argue that listening practices are critical in constructing and maintaining race through language-related practices. Based on this, leading raciolinguists Flores and Rosa (2015) derived the "White listening subject" (p. 151) concept as a linguistic parallel to the "White gaze" concept originally put forward by Morrison in 1998 (Paris & Alim, 2014). White listening subjects listen to the speech of racialized speaking subjects with 'ears' that privilege dominant perspectives and understandings of what racialized speakers say

rather than their physical production of speech (Flores & Rosa, 2015). The White listening subject, rather than being any one physical person or group, is an ideological position such that any person (of any racial background) who adopts the position, can enact Whiteness and White supremacy against racialized speakers through dominant, hegemonic practices of listening (Flores & Rosa, 2015).

Similar to Flores and Rosa, Inoue (2006) philosophizes the importance of broader sociopolitical contexts which shape how gendered speech is categorized and assessed separate from speakers' objective production of speech. She decentralizes the linguistic practices of gendered speaking subjects and centers her analysis around dominant language ideologies associated with social categories that shape listeners' perceptions of gendered speakers and their linguistic practices (Rosa & Flores, 2017). In her study of Japanese 'women's language', instead of drawing attention to and documenting the linguistic practices of Japanese women, such as the features and linguistic variables characteristic of their speech style, she instead focuses on listening subjects' judgments of Japanese women's language and historicizes these assessments within the framework of Japanese politics and economics (Inoue, 2006). She calls this process *indexical inversion* and argues that it plays an active role in constructing how people understand 'reality' in Japan (Inoue, 2006). Drawing from Inoue and her work on indexical inversion, Rosa and Flores (2017) refuse to center their analyses around the linguistic practices of racialized speakers, opting to, instead, problematize practices of White listening subjects who perceive and categorize racialized language in dominant White ways—ways which are themselves perpetuated through socialization.

Combined, experimental and philosophical approaches convincingly demonstrate that, when listeners have additional information about speakers and are privy to the social contexts

within which speakers are situated, non-production-based factors shape listeners' perception and influence the ways in which they categorize and assess speakers and speech. However, when researchers have conducted studies where listeners' categorization and assessments are based on speech alone, the research findings centralized speakers' production of speech and did not devote the same amount of attention to understand the factors shaping listeners' perception (e.g., Baugh, 2000; Gaither et al., 2015; Kurinec & Weaver III, 2019; Idsardi, Purnell & Baugh, 1999; Rickford & King, 2016; Thomas & Reaser, 2004). This is understandable considering that much of the raciolinguistic and psycholinguistic research drawing out the importance of listeners' perception was conducted after the foundational studies on racial identification by speech had already been published (with the exception of Rubin, 1992). Therefore, at the present time, sociolinguists have an opportunity to interrogate the role of listeners' perception when they are hearing disembodied voices and do not have additional information with which to decipher speakers' identities. One way to do this, I propose, is to operationalize listeners' perceptual categorizations and assessments of speakers/speech. Then, the measures can be used in statistical tests to determine whether and to what extent production-based variables (e.g., use of linguistic features) and/or non-production-based variables (e.g., speaker and listener characteristics) predict listeners' categorizations and assessments of speakers/speech. In this chapter, I operationalize listeners' judgments and then test whether these are predicted by speakers' and listeners' demographics (i.e., non-production-based variables). In the following two chapters, I test production-based variables.

4.3 The Need for Listeners' Categorization Measures

Measuring listeners' categorization of speakers and speech and operationalizing these to build a perception-based heuristic can help to improve sociolinguists understanding of how and why listeners categorize speakers the ways they do and illuminate the effects of these categorizations on speakers' social outcomes. Such a measure could, for example, capture the effect of teachers' categorization on students' academic achievement, jurors' categorization on verdict decision-making, 911 operators' categorization on victims' needs, doctors' categorization on patients' diagnoses, and employers' categorization on applicants' qualifications—to name a few. To construct a heuristic to capture these effects, I drew heavily from sociophonetic research on listeners' perception, which highlights some of the (linguistic and non-linguistic) strategies listeners employ to categorize speech and speakers (Clopper, 2004; Foulkes & Docherty, 2006; Foulkes, Scobbie & Watt, 2010). This work centralizes the importance of listeners' own experiences and social knowledge of both language (e.g., language ideologies, linguistic forms) and society (e.g., social groups, stereotypes, ideologies) in determining how and why they categorize speech and speakers in the ways that they do (Clopper, 2021; Levon, 2014). For example, listeners draw from their own social (linguistic and non-linguistic) experiences to form "perceptual representations" by which they compare new stimuli to interpret, assess, and categorize it (Clopper, 2021, p. 337). In turn, each listener's categorization approach differs, and multiple listeners' categorization of speakers are needed to provide a more complete picture of how social processes interact with linguistic practices to shape the ways speakers and speech are interpreted, assessed, and categorized (Clopper, 2021). Based on this, for a heuristic device to be useful across various types of investigations and social classifications, it must contain three properties, detailed below.

4.3.1 Properties of Listeners' Categorization Measures

Operationalizing listeners' categorization of speakers/speech requires: 1) naturalistic speech, 2) perceptual judgments from multiple listeners, and 3) continuous measurements.

4.3.1.1 Naturalistic Speech

Listeners' evaluations of speech and speakers are influenced by their own experiences, which in turn, shape how they evaluate speech *content* (Campbell-Kibler, 2009; 2010). For that reason, in perception experiments, researchers often account for this by using scripted speech to control for the stimuli's content effects (Campbell-Kibler, 2010). However, most instances of non-experimental speech are naturalistic, and research shows that listeners can easily detect the difference between read and naturalistic speech (Blaauw, 1994; Laan, 1997). Thus, using read speech in sociolinguistic experiments on perception threatens the ecological validity of the research study, as it is further removed from the real-world contexts and interactions it is intended to reflect (Campbell-Kibler, 2010, p. 380). In turn, to accurately reflect these contexts and interactions, operationalized measures need listeners' categorizations of naturalistic speech. While naturalistic usually means "casual," what's natural for a particular occasion may actually be more formal—such as language in a courtroom. In addition, naturalistic speech can be captured spontaneously, collected via sociolinguistic interviews, prompted via elicitation tasks, or extracted from the internet, for example (Boyd et al., 2015). Simply put, naturalistic speech can be collected in many settings and in a number of ways (Boyd et al., 2015), but when it comes to operationalizing listeners' categorizations of speech and speakers, the speech that is being used to capture these categorizations must be naturalistic (Simpson, 2013).

4.3.1.2 Perceptual Judgments from Multiple Listeners

Each listener's unique linguistic and non-linguistic backgrounds shape how they categorize speech into socially constructed groups (Fox & Gay, 1994; Kang & Rubin, 2014). Accordingly, listeners with different experiences and backgrounds do not always categorize speech and speakers in the same ways. Taking this into account, to operationalize a perception-based measure and investigate the ways in which listeners' perceptions shape social outcomes, judgments from multiple listeners must be captured. By doing so, the measure becomes capable of more accurately reflecting broader social views about what it means to sound "Black," "gay," or "like a woman," for instance, and is therefore better situated to illuminate how listeners' categorizations influence social outcomes than a single judgment from one listener. Practically, multiple listeners' judgments can be enumerated by averaging their categorizations along one dimension–for example, "Blackness," "gayness," or "femininity."

4.3.1.3 Continuous Measurements

Listeners' perception of speech is profoundly affected by stereotypes and listeners' ideologies (Flores & Rosa, 2015; Levon, 2014). However, perception is also influenced by the extent to which individuals are assessed as fitting perceptual representations (Clopper, 2021). Put another way, even within a particular social category (i.e., "intra-group"), individuals are judged differently based on how much they are perceived as exhibiting stereotypical characteristics. Thus, when operationalizing a perceptual measure to investigate how listeners' categorizations affect social outcomes, it is necessary to take perceived intra-group variation into account. This can be accomplished by operationalizing the measure continuously rather than categorically. In other words, by considering the *likelihood* a speaker or sample of speech will be judged as fitting into a type instead of simply categorizing the speaker or sample of speech into a type (e.g.,

"Black," "gay," "feminine"). Like capturing perceptual judgments from multiple speakers, perceptual measures can be operationalized continuously by averaging speakers' categorizations along one dimension. Therefore, the measure takes into account the odds a listener will categorize a speech sample as sounding like a particular social category.

4.4 Listener-Categorized Speech

Operationalizing listeners' categorizations to meet the requirements outlined above can be accomplished in three steps. The first step is to collect naturalistic speech samples (see Chapter 3). The second step is to recruit listeners to categorize the speech samples. And the third step is to average multiple listeners' categorizations to create a continuous categorization score for each speaker (or speech sample). Taken together, this process allows researchers to construct a perception-based heuristic that captures the likelihood a speaker is assessed by listeners as fitting a particular social category—where ***0*** means the speaker is not at all likely to fit the category, ***0.5*** means the speaker is as likely to fit the category as to not fit the category, and ***1.0*** means the speaker is absolutely likely to fit category. To illustrate this process, I used it to capture listeners' categorizations of "Black-sounding" speech—or *Black racialized speech*.

4.5 Test Case: Black Racialized Speech

To derive and implement a listener-categorized measure, I captured listeners' categorizations of racialized speech. Specifically, I measured the likelihood a speaker was

categorized as "sounding Black" to listeners—a heuristic I call *Black racialized speech*. Accordingly, each speaker was given a Black racialization score which is simply the proportion of times their speech was racialized as "sounding Black" out of all times their speech was racialized by listeners. I selected 200 (PBJ4 and 1-20) of the previously collected naturalistic speech samples to obtain and operationalize listeners' racialization of speakers. I chose to use these as they are analyzed in Chapters 5 and 6 and used in an experiment in Chapter 7. The process I used to obtain and operationalize listeners' racialization of speakers is outlined in the next section.

4.6 Data and Methods

To operationalize the likelihood speakers will be categorized as "sounding Black" by listeners–in other words, to calculate speakers' Black racialization scores–I followed the three steps outlined earlier. First, I collected naturalistic speech samples. Second, I recruited multiple listeners to categorize the speakers. And third, I averaged speakers' categorizations (from multiple listeners) to calculate the final Black racialization score for each speaker.

4.6.1 Data Collection

The naturalistic speech samples I used were collected using the Online Survey Speech Sampling method I outlined in Chapter 3. For this study, I used 200 of the 500 speech samples–the PBJ4 and 1-20 speech samples. My rationale for selecting these samples is explained in Chapter 3.

4.6.2 Perceptual Judgments from Multiple Listeners

I captured perceptual judgments from multiple listeners by recruiting a representative sample of listeners who completed an online survey to categorize the speakers into racialized groups.

4.6.2.1 Listener Recruitment

I recruited a representative sample of 300 listeners using *Prolific.co*–a participant recruitment site that hosts thousands of vetted research participants to complete online studies.

4.6.2.2 Qualtrics Survey

To capture listeners' categorizations, I distributed an online Qualtrics survey embedded with speech samples to the listeners. To complete the survey, each of the 300 listeners listened to 10 speech samples, one at a time, and indicated the racial identity they associated with the speaker in each clip—a process I call *voice racialization*. The listeners could choose only one racial identity for each speech sample: Asian or Asian American, Black or African American, Hawaiian or Pacific Islander, Indigenous American or Alaskan Native, Latinx, Middle Eastern or North African, Multiracial, White or European American. If they selected Multiracial, they were prompted to enter more than one race. While each listener voice-racialized 10 speech samples, I set up Qualtrics randomization so that each listener's set of 10 samples was unique. I also used Qualtrics display logic to ensure that no listener voice racialized a specific speech sample more than once. Additional details on how I created the Qualtrics survey are available in Appendix B.

4.6.2.3 Black Racialization Scoring

To determine the Black racialization score for each speaker, I calculated the proportion of times each speakers' speech sample was voice racialized as "Black" by the listeners who assessed it. In turn, speech samples' Black racialization scores measure the likelihood a speaker "sounds Black" to listeners.

$$\text{Black Racialization Score} = \frac{\text{Number of Black Racializations}}{\text{Total Number of Racializations}}$$

4.7 Analysis

To test the method's effectiveness, I gauged the perceived Blackness of 100 speakers (see Chapter 3 for additional details on the speakers) by assessing listeners' categorization of speakers into their corresponding racial groups. I did this by evaluating how often Black speakers were racialized as "sounding Black," and all other speakers were racialized as sounding like any other race except "Black" by the listeners. Thus, I classified listeners' racialized voice assessments into two categories—Black and Non-Black.[6] If a listener voice-racialized a speaker as 'Black,' then I categorized that racialization as *Black*. Conversely, if a listener voice-racialized a speaker as any other race than 'Black' (e.g., White, Latinx, Asian, etc.), I categorized that racialization as *Non-Black*. Therefore, a Black speaker "sounded Black" to a listener when that listener voice-racialized the speech sample as 'Black.' And a Non-Black speaker did not "sound Black" to a listener when that listener voice-racialized the speech sample as "Non-Black."

[6] I chose to use Black and Non-Black racialized categories for the speakers to mirror the Black racialized speech measure. My rationale for this is explained further in Chapter 7 and I acknowledge and discuss limitations of categorizing speakers this way in Chapter 8.

Using R, I ran descriptive statistics to determine the proportion of times Black speakers were voice-racialized as "sounding Black." To compare, I also determined the proportion of times Non-Black speakers were voice racialized as sounding Non-Black. Additionally, I ran a series of multiple linear regression models predicting whether listeners thought that Black speakers "sounded Black" and Non-Black speakers did not "sound Black." In the models, I included the following control variables: speaker characteristics (race [Black or Non-Black], age, gender, socioeconomic status, and U.S. division), listener characteristics (racial identification, age, gender, socioeconomic status, and whether or not the listener lived in the same region as the speaker), and speech sample type (PBJ4 or 1-20).

4.8 Results

Data collection was very efficient, I distributed the Qualtrics survey on August 1, 2022, at 5:12am EST and received all 300 completed surveys by 4:59pm EST on the same day. The sample of 300 listeners who participated in the study is mostly a representative U.S. sample. However, Latinx listeners are underrepresented. The racial identities, ages, genders, and socioeconomic statuses of the sample of listeners are summarized in Table 4 on the following page.

Altogether, 300 listeners conducted 3,000 voice racialization assessments across the 200 speech samples. Thus, on average, each speech sample was voice-racialized 15 times and each speaker 30 times. The Black racialization scores for all 100 speakers ranged from 0.00 to 0.97 (i.e., no speaker was evaluated as "sounding Black" 100 percent of the time), and the overall mean Black racialization score was 0.30, with a standard deviation of 0.30. The mean racialization score for Black speakers was 0.51, with a standard deviation of 0.28; whereas the mean racialization

score for Non-Black speakers was 0.08, with a standard deviation of 0.09. The sample means are different and are statistically significant ($p < 0.001$).

Table 4 Descriptive Summary of Listeners' Demographics

	Mean	Standard Deviation	Min	Max
Listener Characteristics				
Racial Identification				
White	0.72	0.45	0	1
Black	0.14	0.35	0	1
Latinx	0.04	0.20	0	1
Asian	0.07	0.25	0	1
Indigenous	0.02	0.14	0	1
Middle Eastern	0.00	0.06	0	1
Age	45.85	15.70	18	81
Gender				
Man	0.48	0.50	0	1
Transgender Man	0.01	0.08	0	1
Non-Binary	0.00	0.06	0	1
Transgender Woman	0.00	0.06	0	1
Woman	0.51	0.50	0	1
Socioeconomic Status	2.97	0.92	1	5
Number of Listeners (Speakers)		300 (100)		

Roughly half or 1,508 of the 3,000 voice racialization assessments were of Black speakers' samples and the remainder (1,492) were of Non-Black speakers' samples. Of the 1,508 voice racialization assessments for Black speakers, 766 were racialized as "sounding Black." Put another way, listeners thought that Black speakers "sounded Black" 50.82 percent of the time. Conversely, of the 1,492 voice racialization assessments for Non-Black speakers, 1,368 were not racialized as "sounding Black." In other words, listeners thought that Non-Black speakers sounded "non-Black" about 91.67 percent of the time. Altogether, listeners thought that Black speakers' speech samples sounded Black and Non-Black speakers' speech samples did not sound Black 71.14 percent of the time (see Figures 1 and 2).

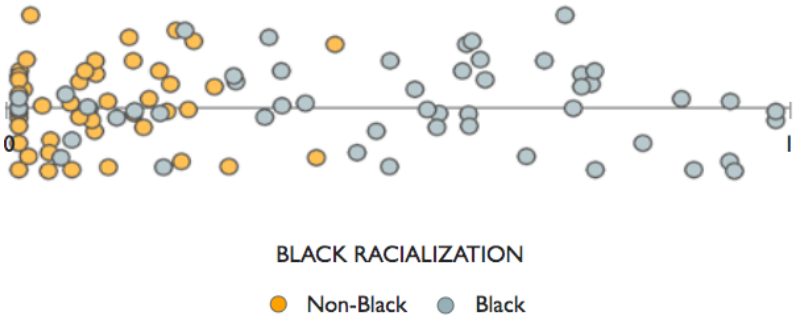

Figure 1 Black Racialization Scores for 100 Speakers (Scatterplot)

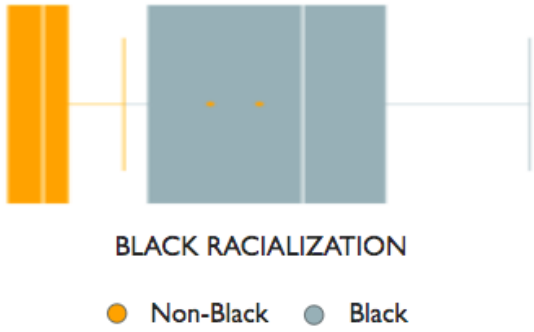

Figure 2 Black Racialization Scores for 100 Speakers (Boxplot)

The regression tests revealed overall, listeners were significantly more likely to say that Non-Black speakers did not sound Black than to say that Black speakers sounded Black. Few demographic characteristics factored into this. However, when speakers were from the Eastern South (AL, KY, MS, TN) and Western South (TX, OK, AR, LA) divisions, listeners were more likely to say that Black speakers sounded Black and that Non-Black speakers did not sound Black compared to speakers from other divisions. In other words, when listeners heard speakers from southern U.S. regions, they were more likely to perceive and racialize speakers as sounding like their corresponding self-identified racial identities (on a Black-Non-Black binary) compared to when they heard speakers from other regions. Plainly, listeners were better at accurately identifying speakers' racial identities when the speakers were from the south.

Table 5 Coefficients from Voice Racialization Linear Regression Models

	Model 1	Model 2	Model 3
		Mean (St. Err.)	
Speaker Characteristics			
Black	-2.42 (0.11)*	-2.51 (0.13)*	-2.52 (0.13)*
Age		0.00 (0.00)	0.00 (0.00)
Gender (Ref. Man)			
Transgender Man		-0.76 (0.41)	-0.74 (0.42)
Woman		-0.03 (0.10)	-0.02 (0.10)
Socioeconomic Status		-0.00 (0.06)	-0.01 (0.06)
U.S. Division (Ref. New England)			
Middle Atlantic (NY, NJ, PA)		0.31 (0.24)	0.30 (0.24)
Eastern Midwest		0.34 (0.23)	0.31 (0.23)
Western Midwest		-0.04 (0.26)	-0.04 (0.26)
Eastern South		0.51 (0.21)*	0.47 (0.21)*
Deep South (KY, TN, MS, AL)		0.41 (0.25)	0.40 (0.25)
Western South (TX, OK, AR, LA)		0.79 (0.25)*	0.79 (0.25)*
Mountain		0.24 (0.28)	0.25 (0.28)
Pacific (WA, OR, CA, HI, AK)		-0.11 (0.25)	-0.12 (0.25)
Listener Characteristics			
Racial Identification (Ref. White)			
Black			0.15 (0.13)
Latinx			0.09 (0.23)
Asian			0.12 (0.19)
Indigenous			-0.57 (0.31)
Middle Eastern			-1.46 (0.56)*
Age			-0.00 (0.00)
Gender (Ref. Man)			
Transgender Man			0.24 (0.61)
Non-Binary			-0.13 (0.69)
Transgender Woman			-0.68 (0.85)
Woman			0.00 (0.09)
Socioeconomic Status			-0.03 (0.05)
Lives in Same Division as Speaker			0.07 (0.13)
Natural Speech Clip (PB&J)	0.02 (0.09)	0.01 (0.09)	0.02 (0.09)
Constant	2.43 (0.11)*	2.06 (0.33)*	2.30 (0.40)*
Pseudo R^2	0.1893	0.1990	0.2030
Number of Clips Categorized		3000	

* Denotes that the coefficient is statistically significantly different from zero with a p-value < 0.05

4.9 Discussion

Currently, sociolinguists have an opportunity to investigate the role of listeners' perception in speaker/speech categorization processes even when these listeners do not have additional information about speakers' identities. I operationalized listeners' perceptual categorization of speakers as "sounding Black" to begin to tease apart perception from production in speaker/speech categorization of disembodied speech (i.e., when listeners do not know or cannot decipher information about speakers' identities). Doing this required completing three steps: collecting naturalistic speech samples, recruiting multiple listeners to categorize speakers into racialized groups, and averaging listeners' categorizations to calculate the Black racialization score for each speaker. I used multiple linear regression models with and without control variables to gauge the perceived Blackness of 100 speakers using a Black-Non-Black binary for race.

Overall, listeners were more likely to categorize Non-Black speakers as "sounding Non-Black" than to categorize Black speakers as "sounding Black." This is reasonable considering that "Non-Black" is a much broader category than "Black." For example, four different listeners could have categorized the same speaker as "sounding White", "Asian", "Latinx", or "Indigenous" and all of these would have fallen within the scope of "sounding Non-Black." Despite this, speakers' likelihood of being categorized into a racial identity group that matched their own (on a Black-Non-Black binary) was much greater than chance. This is the case even though listeners were not informed of speakers' racial identities, nor were they given external physical (or other) cues to decipher speakers' racial identities. In other words, they categorized speakers based on listening alone.

4.10 Conclusion

In this chapter, I aimed to demonstrate the need for perceptual measures of listeners' categorization of speakers in sociolinguistic research. Specifically, in doing so, I began to justify my choice to use Black racialized speech, which is based on listeners' perception, and to shift from focusing on production in this research. Making this shift has important ramifications for addressing my overarching research question. And because Black racialized speech is a continuous variable, I will be able to test whether a higher Black racialization score—or being more likely to be categorized as "sounding Black"— more strongly correlates with criminality compared to a lower Black racialization score. In the chapters immediately following this one, I provide further support for using Black racialized speech in this research, relying on empirical evidence instead of theory to support my proposition.

5.0 Sounding Black: Speakers' Use of AAVE Features and Listeners' Voice-Racialization

5.1 Introduction

African American Vernacular English (AAVE) is one of the most widely studied dialects of American English to date (Farrington, King & Kohn, 2021; King, 2020). From the variety's phonology and morphology to its grammar and lexicon, well-documented features like habitual *be* and copula deletion have, since the 1960s, characterized both AAVE and sociolinguists' ideas about what it means to "sound Black" in America (e.g., Fasold, 1969; Labov, 1969; Wolfram, 1969). The past half-century's work on AAVE has demonstrated that Black speakers' language is systematic, grammatical, and predicable, like all other language varieties (Green, 2002; Lanehart, 2015; Smitherman, 1988). Early scholars set out to document this systematicity in order to address negative ideologies about AAVE that contributed to persistent racial inequities in schools and to earn important educational rights for Black students (King, 2020; Labov, 1968; Smitherman, 1996; 1999). Later on, sociolinguists similarly set out to address how these ideologies contributed to racial inequities in the *criminal legal system* and negatively affected outcomes for Black defendants and witnesses (Corley [Branson], 2014; Jones et al., 2019; Kurinec & Weaver III, 2019; Rickford & King, 2016). In fact, a significant proportion of published research on the relationship between "sounding Black" and criminality is on AAVE.[7]

However, when some researchers have tried to isolate the AAVE features that correlate with "sounding Black," their findings have been inconclusive (e.g., Gaither et al., 2015; Thomas

[7] Jones et al.'s 2019 study uses "African American English (AAE)." However, they explain on p. e216 that they use "AAE" as an alternate for AAVE, "among others."

& Reaser, 2004). This suggests that, even when listeners hear disembodied voices, AAVE use may not be the only cue they rely on to decipher that a speaker is Black. Accordingly, it may not be the only variable affecting listeners' criminalization of "Black-sounding" speakers. For example, Weissler (2022), whose own research engages psychology, linguistics, and Black studies, argues that listeners' ideologies (e.g., stereotypes) mitigate their perceptions of, and how they assess, Black speakers' language. This similarly implies that use of AAVE production features alone does not fully capture what it means to "sound Black" in the United States, nor does it fully explain the criminalization of "Black-sounding" speakers.

Black racialized speech, which is derived from listeners' perception, directly represents speakers' likelihood of being categorized as "sounding Black" by listeners. Accordingly, it may encompass more ways to "sound Black" than are represented by one racialized dialect—AAVE. Yet, while I theorized and supported this argument with psycholinguistic and raciolinguistic literature in Chapter 4 (Eberhardt et al., 2004; Eberhardt et al., 2006; Flores & Rosa, 2015; Hay & Drager, 2010; Levon, 2006; Niedzielski, 1999), I have not yet empirically demonstrated whether and to what extent listeners' judgments are based on their own perceptual biases (or, potentially other factors), rather than speakers' production, of speech. One way to do this is to tease apart listeners' perception of speakers' racial identities from speakers' production of AAVE. I applied this approach to test whether speakers' use of AAVE production features predicted listeners' categorization of the speakers as "sounding Black."

To do this, I systematically evaluated the extent to which each of the 100 speakers in this study used six different AAVE phonological features[8]. I conducted multiple linear regression tests

[8] I focused on phonological features only because an initial review of the data revealed that speakers did not use AAVE morpho-syntactic features and lexical items. This is likely because of the limited context within which speakers created the speech samples.

and found, as I hypothesized, that use of these features did not predict whether listeners thought speakers "sounded Black." In what follows, I describe my process of evaluating the speech samples and conducting the statistical tests. After, I discuss the implications of these findings, and begin to build an empirical case to focus more on listeners' perception rather than speakers' production of speech in research on Blackness, language, and criminality.

5.2 African American Vernacular English Features

While there are many overlapping features used in AAVE and standardized American English (sdAE), the varieties have differences at every level of linguistics (Green, 2002). These areas of difference are where linguists have drawn out production features to characterize AAVE (King, 2020). In fact, some of AAVE's most characterizing features—like invariant *be* and remote past BIN—have no simple equivalent in sdAE (Green, 2002). Below, I overview AAVE production features that sociolinguists have used to characterize the variety over the years, including features from AAVE's lexicon, grammar, and phonology. Later, I list the AAVE features I chose for the empirical tests and explain why I chose those features.

5.2.1 Lexicon

AAVE's lexicon is like a dictionary, or a repository for the words that make up the language variety (Green, 2002). In many cases, AAVE's lexicon overlaps with sdAE's, yet AAVE's lexicon also includes many words that are unique to the variety as well as words that are also in sdAE but have different meanings in AAVE. *Ashy* and *saditty,* for instance, are unique

AAVE words. Ashy refers to "the whitish coloration of black skin due to exposure to…cold and wind" (Smitherman, 1977, p. 67) and saditty refers to "uppity-acting blacks who put on airs" (Smitherman, 1977, p. 68). *Kitchen,* on the other hand, is also an sdAE word but has another meaning in AAVE. In AAVE, kitchen also means "the hair at the nape of the neck which is inclined to be very kinky" (Smitherman, 1977, p. 64). 'Ashy', 'saditty', and 'kitchen' have been used across generations of AAVE speakers, evidencing that AAVE's lexicon is not simply a bank of slang words, though it is sometimes portrayed this way in broader society. Yet, like all language varieties, AAVE does have its slang—words like, "*benjis*, *cabbage*, *cheese*, *cream*, *duckets*, *franklins*, *paper*, and *scrilla*," which all mean "money" (Green, 2002, p 29-30).

5.2.2 Grammar

AAVE's grammar refers to the variety's unique morphosyntactic production features and patterns (Kendall & Farrington, 2021). Early studies of Black speakers' language primarily focused on AAVE's grammar (e.g., Baugh, 1979; Fasold, 1972; Labov, 1968; Wolfram, 1969). Therefore, much of what sociolinguists know about AAVE has to do with its morphosyntactic features—particularly the features that characterize AAVE's *tense*, *aspect*, and *mood* system (Kendall & Farrington, 2021). Invariant *be* and remote past BIN, for example, are two of AAVE's most iconic tense-aspect features. Invariant *be* indicates that a described event takes place across time or space (Green, 2002; Kendall & Farrington, 2022). Put another way, it describes an event that occurs habitually (Labov et al., 1968; Fasold, 1972; Dayton, 1996; Green, 2002). So, saying, "I *be* at the store," in AAVE means that I am usually at the store, but it does not and cannot mean that I am currently at the store. Remote past BIN indicates that an action or event happened in the past and was completed earlier on (Spears, 1982; Wolfram, 2004). So, "I BIN ate," means that I ate

and finished eating a while ago. Another unique feature of AAVE morphosyntax is that the variety does not require use of the copula *is* in every instance where it is required in sdAE—what some sociolinguists call *copula deletion* (Green, 2002). For example, I could say in AAVE, "He nice," to mean, "He *is* nice" in sdAE. In addition, AAVE, like some other non-standardized American English dialects, exhibits negative concord (Green, 2002; Wolfram, 1969; 2004), or making a single negative proposition with multiple negative markers—for example, "*Ain't nobody* got time for that" to mean "Nobody has time for that" in sdAE. AAVE is rich in unique morphosyntactic features; however, documenting them all is far beyond the scope of this dissertation.

5.2.3 Phonology

AAVE's phonological system refers to the collection of sounds that make up AAVE as well as the rules dictating how these sounds interact with one another (Kendall & Farrington, 2021). Like AAVE's grammatical system, its phonological system has similarly unique features. Since describing AAVE is not my dissertation's primary purpose, I will limit this section to describing the six phonological features I ultimately selected for the analysis.

A number of AAVE phonological features are characterized by rules applied to consonant sounds at the ends of words (Thomas, 2015). For example, when sdAE words end in constant clusters (i.e., more than one consonant), in AAVE the second (or final) consonant is sometimes not present—like saying *kin'* for *kind* (Baugh, 1983; Labov et al., 1968; Thomas, 2007; Wolfram, 1969; Wolfram & Thomas, 2002). Sociolinguists call this process *consonant cluster reduction*. Individual consonants at the ends of words are sometimes not present in AAVE speech, too (Moran, 1993)—like when saying [æ̃] for *at*. Similarly, velar nasal fronting, also called *g-dropping*, occurs in AAVE (Labov et al., 1968; Forrest & Wolfram, 2019). And, when sdAE words

end in *-th*, AAVE speakers may pronounce them with an /f/ or /v/ instead (Green, 2002). Conversely, if *th-* is at the beginning, AAVE speakers may use /t/ or /d/ (Green, 2002). In unstressed syllables, some AAVE speakers may not pronounce an *r* sound that would otherwise be pronounced in sdAE, like saying *motha* for *mother* (Hinton & Pollock, 2000; Labov et al., 1968: McLarty, 2019). And some AAVE speakers are known to merge the vowels in *pen* and *pin* before nasal sounds (Charity, 2008). In other words, they may say *pin* for both *pin* and *pen.*

Taken together, there are many lexical, grammatical, and phonological features that set AAVE apart from sdAE. Therefore, sociolinguists have reasonably assumed that when Black speakers use AAVE production features listeners categorize them as "sounding Black" because of it (Baugh, 2000; Gaither et al., 2015; Purnell, Idsardi & Baugh, 1999). However, I have chosen to use Black racialized speech rather than AAVE in my research because I have theorized that listeners' *perceptual* biases—and not only because of the ways in which Black speakers *produce* speech—shape how listeners 'hear', evaluate, racialize, and possibly criminalize speech. To empirically justify my decision to use Black racialized speech rather than AAVE, I tested whether speakers' use of AAVE production features predicted how listeners voice racialized their speech.

5.3 Methods

To test whether speakers' use of AAVE production features predicted the likelihood listeners categorized them as "sounding Black," I systematically scored each of the PBJ4 speech samples (collected previously) based on speakers' use of the six AAVE phonological features outlined in the previous section. I describe this process in detail below.

5.3.1 AAVE Scoring Process (PBJ4 Speech Samples)

I began the AAVE scoring process by making hand corrections to the automated transcripts provided by Phonic. I converted each individual transcript to an Excel file, where I conducted the scoring. I scored the samples based on the extent to which each sample exhibited the following AAVE phonological features: consonant cluster reduction, *th* → *t/d* or *f/v*, *r*-lessness, velar nasal fronting, *pin-pen* merging, and final consonant deletion. I chose these features after my initial review of the transcripts revealed that speakers did not use unique lexical or grammatical AAVE production features. This is likely due to the limited context in which the speech samples were collected (see Chapter 3). Nevertheless, I noted that speakers varied in their use of these six AAVE phonological features, which I selected for the analysis (see Table 6).

Table 6 AAVE Phonological Features

Feature	Definition	Example
Consonant Cluster Reduction	"A process in which the final consonant group or cluster, composed of two consonant sounds, is reduced to a single consonant sound" (Green, 2002, p. 107).	CC → C || | nd → d kind→kin
th → *t/d* Word Initially *th* → *f/v* Word Finally	"The production of *t*/*d*…in environments in words in which the *th* sound occurs in general American English" (Green, 2002, p. 117).	th → d these → dese th → t month → monf
r-lessness in Unstressed Syllables	"Deletion of the /r/ sound following a vowel or between two vowels" (Kendall & Farrington, 2021)).	father → fatha
Velar Nasal Fronting	"Common feature of all varieties of American English, where *-ing* is replaced by *-in*" (Kendall & Farrington, 2021).	coming → comin
Pen/*Pin* Merger	"The merger of the vowels [ɛ] and [ɪ] before nasals" (Charity, 2008, p. 36)	pen → pin
Final Consonant Deletion	Final consonant in a word is deleted, the preceding vowel is usually lengthened (Moran, 1993).	at → ǣ

I calculated an AAVE score for each speech sample by identifying and counting all instances in which the AAVE phonological rules were applied by the speaker in the sample. Then, I divided that number by the total number of instances in which the AAVE phonological rules *could have* been applied. I used my own native speaker knowledge of AAVE and sdAE to determine when and where the phonological rules could have been applied and listened to each speech sample slowly and carefully to determine if the rules were applied. So, in the end, each speech sample's AAVE score represented the proportion of times I heard the speaker apply the AAVE phonological rules out of all times when they could have applied the rules.

5.4 Analysis

To analyze the data and test whether the AAVE scores correlated with speakers' Black racialization scores (calculated in Chapter 4), I ran descriptive statistics and two multiple linear regression models using *R*. The descriptive statistics simply provided the mean, standard deviation, and range of the AAVE scores. I also transformed the speakers' scores to graphs and charts for visual comparison. Concerning the regression tests, both models predicted the speech samples' Black racialization scores (dependent variable). In Model 1, speakers' AAVE scores was the key independent variable and speakers' race (Black or Non-Black) was the control variable. In Model 2, AAVE scores remained the key independent variable and speakers' race was still a control variable, but I also included the following additional control variables in Model 2: age, gender, socioeconomic status, and U.S. division.

5.5 Results

Across all 100 PBJ4 speech samples, the mean AAVE score was 0.19, with a standard deviation of 0.14. The lowest AAVE score was 0.00 and the highest was 0.67. The scatterplot and boxplot below indicate a fair amount of overlap between Black and Non-Black speakers' PBJ4 AAVE scores and all speakers, except for two, used AAVE features less than half of the time. Plainly, speakers mostly used sdAE features regardless of their race.

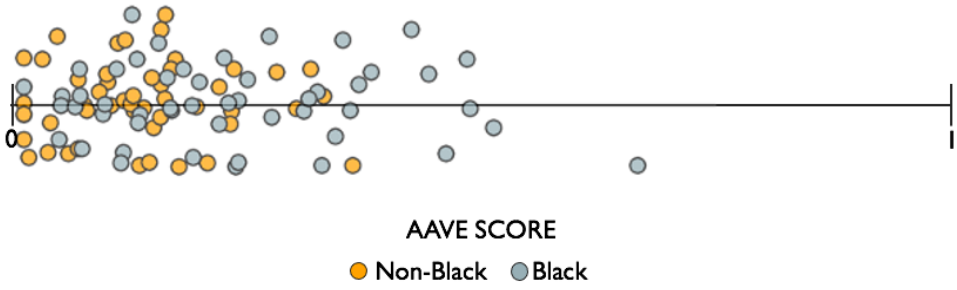

Figure 3 AAVE Scores for PBJ4 Speech Samples (Scatterplot)

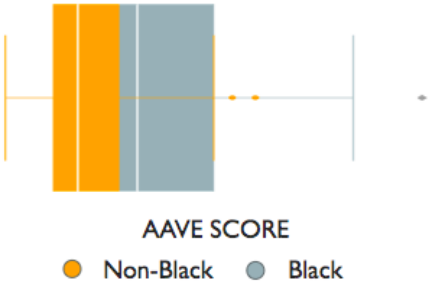

Figure 4 AAVE Scores for 100 PBJ4 Speech Samples (Boxplot)

The results of the regression tests revealed that speakers' AAVE scores—or the extent to which they used the six AAVE phonological features in their PBJ4 speech samples—postively correlated with being categorized as "sounding Black." In other words, using more AAVE phonological features increased the likelihood a speaker "sounded Black" to listeners. However, the AAVE scores were *not* statistically significant predictors of this likelihood. In addition, despite

the fact that listeners did not explicitly know speakers' racial identities nor have other external cues to decipher speakers' racial identities, the control variable, *race* (Black or Non-Black), was statistically significant—with Black speakers having an increased likelihood of "sounding Black" to listeners. In Model 2, I included additional speaker characteristics as control variables and *age* was also positively correlated with speakers' Black racialization scores and statistically significant, meaning that older speakers were more likely to "sound Black" to listeners. Otherwise, none of the remaining speaker characteristics predicted the voice racializations. (See Table 7 for complete list of speaker characteristics included in the models.)

Table 7 Coefficients from Regression Models Predicting Proportion of Clips Racialized as 'Black'

	Model 1	*Model 2*
AAVE Score	0.34 (0.17)	0.37 (0.19)
Speaker Characteristics		
Black	0.39 (0.05)*	0.36 (0.05)*
Age		0.00 (0.00)*
Gender (Ref. Man)		
Transgender Man		-0.33 (0.23)
Woman		0.00 (0.05)
Socioeconomic Status		-0.02 (0.03)
U.S. Division (Ref. New England)		
Middle Atlantic (NY, NJ, PA)		-0.05 (0.12)
Eastern Midwest		-0.08 (0.11)
Western Midwest		-0.04 (0.12)
Eastern South		-0.01 (0.11)
Deep South (KY, TN, MS, AL)		0.04 (0.14)
Western South (TX, OK, AR, LA)		0.16 (0.12)
Mountain		0.02 (0.14)
Pacific (WA, OR, CA, HI, AK)		-0.06 (0.13)
Constant	0.02 (0.04)	-0.08 (0.16)
R^2	0.4929	0.5950
Number of Clips Categorized	100	

*Denotes that the coefficient is statistically significantly different from zero with a p value < 0.05

5.6 Discussion

In past research on racial categorization by speech, sociolinguists attributed listeners' identification of Black speakers' voices to the speakers' use of AAVE production features (e.g., Idsardi, Purnell & Baugh, 1999). The findings in this chapter, however, do not fully support this argument. In this analysis, across all 100 speakers, use of six AAVE phonological features did not significantly predict the likelihood a speaker "sounded Black" to listeners. This could possibly be due to speakers' infrequent use of AAVE phonological features. In fact, earlier research has shown that listeners are worse at identifying speakers' racial identities when the speakers do not use AAVE features (e.g., Thomas & Reaser, 2004). However, despite speakers' infrequent use of AAVE phonological features, speakers' own racial identities and age were statistically significant predictors of their likelihood of "sounding Black" to listeners. And this was the case even though the listeners did not know speakers' age or race. This suggests that listeners may draw from other, non-dialectal cues to determine speakers' racial identities or perhaps their own perceptual biases are playing a role. For example, listeners' association of Black racialized speech with older speakers may be explained by their perception of the speakers' voice quality rather than use of AAVE features. Taken together, I conclude that the insignificant, but positive, correlation between use of AAVE phonological features and speakers' likelihood of "sounding Black" to listeners (i.e., Black racialization scores) indicates that listeners may use AAVE phonological features as cues to decipher speakers' racial identities, but this is not the only—and possibly not the primary—way they do so. In turn, the findings of this study bolster the case to further investigate other, non-production-based factors that shape listeners' perception, categorization, and assessment of speakers/speech over speakers' production of AAVE.

In the context of my broader research question, *Do U.S. listeners associate Black racialized speech with criminality*, the findings outlined in this chapter suggest that, if listeners do associate Black racialized speech with criminality, it is not solely because the speakers they are listening to use the six AAVE phonological features I tested. Rather, listeners' own perceptual biases are possibly playing a role. By taking this approach to centralize listeners' perception, this work importantly avoids inadvertently attributing Black-sounding speakers' criminalization to their own use of AAVE phonological features. Yet, the positive, albeit insignificant, relationship between use of AAVE features and "sounding Black" indicates that listeners' may be more likely to think AAVE speakers "sound Black." Therefore, the ways in which negative ideologies about AAVE shape the criminalization of Black speakers, as noted in earlier research (e.g., Jones et al., 2019; Kurinec & Weaver III, 2019; Rickford & King, 2016), and are used as an excuse to "justify" anti-Black discrimination, should not be taken lightly or disregarded.

5.7 Conclusion

In this chapter, I aimed to test whether speakers' use of AAVE phonological features predicted whether they "sounded Black" to listeners. Regression tests revealed that speakers' use of these features did not significantly predict this. However, speakers' self-identified race and age were statistically significant control variables explaining listeners' racialization of speakers' voices—and this is despite the fact that listeners did not know this information explicitly or have other cues to decipher speakers' racial identities or age. The findings of this analysis begin to build the case the speakers' production of speech is not the only, or most important, factor shaping how listeners 'hear,' racialize, and assess speech. Accordingly, U.S. listeners' association of Black

racialized speech with criminality, then, would not be fully explained by Black speakers' use of AAVE features. Instead, the findings suggest that listeners own perceptual biases may be factoring in. In the following chapter, I conduct one final production-based analysis of other features that have been documented in the literature as characteristic of speech in Black communities (i.e., African American Language). The goal of the next chapter, Chapter 6, is to further bolster the case that speakers' production of speech is not the primary predictor of listeners' racial categorizations and assessments of speech/speakers.

6.0 Beyond AAVE, Towards AAL: AAL Features and Listeners' Voice Racialization

6.1 Introduction

For over 200 years, researchers have tried to isolate the physical and biological characteristics (e.g., genetics, skull sizes/shapes, skin color) that distinguish humans into racial categories and characterize how each group "looks" (e.g., Blumenbach, 1795; Bolnick, 2008; Guo et al., 2014; Lombroso, 1876; Shiao et al., 2012). Yet, research study after research study has shown that racial identities cannot be delimited to a mere set of physical features or biological characteristics (e.g., Morning, 2011; Morning, 2014; Serre & Pääbo, 2004; Smedley & Smedley, 2005). Rather, racial identities (and what they allegedly "look" like) are undeniably social constructs (Douglass et al., 2016; Kendi, 2023; Lopez, 1994; Omi & Winant, 2014). Sociolinguists have similarly tried to identify vocal production features that distinguish racial groups and characterize what it means to *sound* like a particular race (Gaither et al., 2015; Thomas & Reaser, 2004; Perrachione et al., 2010; Purnell, Idsardi & Baugh, 1999). For example, sociolinguistic researchers have attempted to correlate speakers' use of AAVE features with listeners' categorization of speakers as "sounding Black." And like in biology and anthropology, identifying these features has proved difficult[9] (e.g., Gaither et al., 2016; Thomas & Reaser, 2004; Purnell, Idsardi & Baugh, 1999). Some researchers in different subfields of linguistics, (e.g.

[9] I identified one study (Perrachione et al., 2010) in which speakers' vocal production of AAVE features statistically significantly predicted listeners' racialized assessments of speech (i.e., whether listeners thought the speech sounded Black or White). However, the study included a small sample of speakers (12 Black and 12 White) who were all male and the study's listeners were also a small group (10 Black, 10 White) and predominantly female (F = 17, M = 3). Thus, the results of the study are not generalizable to Black speakers more broadly.

psycholinguistic, sociophonetics, raciolinguistics) have put forth that speakers are assessed as sounding like racial identities not because of the features the use, but because of social practices and psychological processes that shape listeners' perceptions of speech and speakers (Alim et al., 2016; Flores & Rosa, 2015; Rosa & Flores, 2017; Niedzielski, 1999). Thus, "sounding like a race" is also a social construct that cannot be delimited to a set of racialized linguistic features (Flores & Rosa, 2015; Rosa & Flores, 2017).

Like Chapter 5, this chapter aims to empirically demonstrate that the production of speech is not the primary cue listeners rely on to determine speakers' racial identities. However, in contrast to Chapter 5 and earlier studies testing whether AAVE use predicted listeners' categorization of speakers' racial identities (e.g., Gaither et al., 2015; Thomas & Reaser, 2004), in this chapter, I test a broader range of features from language varieties spoken in African American communities (i.e., African American Language, or *AAL*). AAL is an umbrella term referring to all language varieties spoken in African American communities (Kendall & Farrington, 2021; Lanehart, 2015; Lanehart, 2022). Therefore, AAL encompasses more Black speech practices, and presumably more ways to "sound Black" than AAVE, which represents the speech practices of just one subgroup of Black speakers (King, 2020; Lanehart, 2015; Lanehart, 2022). By narrowly focusing on AAVE, earlier research studies may have unintentionally overlooked speech practices that are characteristic of a wider range of Black-sounding speech, which could explain why AAVE vocal production features did not correspond with listeners' assessments that speakers "sounded Black."

To investigate additional "Black-sounding" production features, I analyzed 200 speech samples (100 PBJ4 and 100 1-20) from Black and Non-Black speakers to test three features that earlier studies suggested may distinguish speech in Black communities (i.e., AAL) from speech in White (and other) communities. These are: articulation rate (Kendall, 2013), front-stressing of

initial syllables (Baugh, 1983; Fasold & Wolfram, 1970; Kendall & Farrington, 2021; Sutcliff, 2003; Thomas, 2015), and pronunciation of LOT, THOUGHT, TRAP, DRESS, and KIT vowels (e.g., Farrington, King & Kohn, 2021, Thomas, 2001; 2007). A series of regression tests found no correlation between the majority of these production features and listeners' racialization of speakers (i.e., judging speakers as sounding Black or Non-Black). However, I found that Black speakers' production of the vowel in the word *TEN* (DRESS or KIT vowel depending on pronunciation) was significantly different from Non-Black speakers' production of the vowel, and this corresponded with an increased likelihood that listeners racialized speakers differently. In what follows, I describe AAL and distinguish it from AAVE. Then, I detail the three features I tested, give an overview of the analytical methods, and provide detailed results of the regression analyses. I then discuss the results, with special attention to contextualizing and problematizing the results from the analysis of the vowel in *TEN*. Finally, I discuss the implications of the study's findings to the main research question and conclude.

6.2 Distinguishing AAVE from AAL

Over the past 60 years, sociolinguists have extensively documented speech practices in Black communities throughout the United States (e.g., Baugh, 1979; Fasold, 1972; Labov, 1968; Wolfram, 1969). Early work on this was part of a coordinated effort to address racial disparities in U.S. schools by demonstrating that Black speakers in cities across America spoke a unique, but not broken, variety of American English (King, 2020). The success of this effort rested on sociolinguists' abilities to show that speech in Black communities was systematic, rule-governed, and importantly, homogenous (Labov, 1968; King, 2020). Thus, despite the range of speech styles

represented by Black speakers, sociolinguists mostly focused on one subgroup–young Black men in the inner city—and the language practices of this group came to be known as AAVE (Labov, 1972; Morgan, 1994; Smitherman, 1988; Wolfram, 2007). Since the 1980s, however, other sociolinguists have pointed out that young, urban Black men's speech is not representative of all Black speakers' speech practices and have called for more diverse representation of speakers in research on language in Black communities (Morgan, 1994; Rickford, 1997; Smitherman, 1988). These calls for broader representation resulted in the creation of a body of intersectional research centering other subgroups of Black speakers (e.g., Black women and girls, middle-class African Americans, rural African Americans, Gullah speakers) and a more inclusive approach to studying speech in Black communities–African American Language (AAL) (Cukor-Avila & Bailey, 2015; Lanehart, 2015; Smitherman, 1988; Weldon, 2021; Weldon & Moody, 2015)

As noted above, by looking at AAL, and thus focusing on a broader range of ways of speaking and potentially "sounding Black," I expand the review of Black speech practices beyond just one subgroup. Leading AAL scholar, Sonja Lanehart explains, "...when speakers know AAL, they know a system of sounds, word and sentence structure, meaning and structural organization of vocabulary items, and other linguistic and metalinguistic information about their language, such as pragmatic rules and the social function of AAL" (Lanehart, 2015, p. 2). While much more extensive research has been conducted on AAVE, scholars have identified several features AAL speakers "know" (Lanehart, 2015, p. 2) that may distinguish speech in Black communities (i.e., AAL) from speech in other communities. These features include, articulation rate (Kendall, 2013), front-stressing of initial syllables (Thomas, 2015), and pronunciations of vowels (Farrington, King & Kohn, 2021).

6.3 Some AAL Features

6.3.1 Articulation Rate

Articulation rate is a rate-of-speech measure (Goldman-Eisler, 1961). It differs from "speech rate," as it is calculated using the speaking portions of speech only–pauses are not included (Goldman-Eisler, 1961; Kendall, 2013). Thus, articulation rate is a specific type of rate-of-speech measure that takes into account how fast or slow a person produces, or *articulates*, their speech. Despite widely held ideas that speakers' articulation rates vary by region, gender, race, or age—for example, Southern U.S. speakers are thought to be slow talkers compared to New Englanders—research conflicts as to whether Americans' articulation rates are objectively different across social groups (Clopper & Smiljanic, 2007; Kendall, 2013; Robb, Maclagan & Chan, 2004; Salmons et al., 2008; Yuan et al. 2006). Kendall (2013) tested variation in the articulation rates of 104 speakers from different regions, ethnic groups, and genders. Based on these speakers' production of over 23,000 syllables, Kendall found that White Americans in his study (N = 31) had a higher mean articulation rate (mean = 5.22 σ/sec) than Black Americans (N = 34, mean = 4.46 σ/sec), and this difference was statistically significant ($p < 0.001$). Thus, Black speakers in his study were slower talkers than their White counterparts. Despite this, and widely held beliefs about speech rates across racial groups, cross-racial speech and articulation rate studies are still in the initial stages (Thomas, 2015). In fact, I conducted a comprehensive review of the sociolinguistic literature and was unable to identify another study that systematically compared speech rates across Black and Non-Black racial groups.

6.3.2 Front-Stressing of Initial Syllables

Front-stressing of initial syllables, also called *forestressing*, is the process of shifting words' primary stress from a later syllable to the first one (Baugh, 1983; Kendall & Farrington, 2021; Thomas, 2015). Sociolinguists have documented front-stressing by Black speakers in words like: *PO.lice*, *JU.ly*, *AD.dress*, *DE.troit*, and *HO.tel*, among others (Baugh, 1983; Fasold & Wolfram, 1970; Kendall & Farrington, 2021; Thomas, 2015). In each of these two-syllable words, standardized American English pronunciations place primary stress on the second syllable (Oxford English Dictionary, 2023). With front-stressing, however, primary stress is on the first syllable (Baugh, 1983; Kendall & Farrington, 2023; Thomas, 2015). While southern White speakers are also known to front-stress, research shows that front-stressing is fairly widespread among speakers in Black communities throughout the United States and not just in the south (Pedersen et al., 1986-92 in Thomas, 2015). However, among Black speakers, front-stressing is stereotypically associated with older speakers and those from lower socioeconomic backgrounds (Pedersen et al., 1986-92 in Thomas, 2015).

6.3.3 Vowel Pronunciation

Vowel sounds differ based on a speaker's tongue position (front/back), roundness (rounded/unrounded), and height (high/low) (Ladefoged, 2006). We can observe variation in actual vowel *pronunciation*, based on how the vowel is produced in a particular word, which may be influenced by social factors (MacFarlane & Stuart-Smith, 2012; Thomas, 2007). For example, speakers may pronounce a vowel in a word differently–like saying [pɪn] or [pɛn] for "pen"—based on region (Austen, 2020; Bailey, 1997; Montgomery & Eble, 2004; Thomas, 2020). Sociolinguists

have identified that some speakers in Black American communities utilize a unique vowel system called the African American Vowel System (AAVS) (Farrington, King & Kohn, 2021). Thomas's work was instrumental in identifying and laying out important distinctions between Black and White speakers' pronunciation of LOT, THOUGHT, TRAP, DRESS, and KIT vowels, which characterize the AAVS. The AAVS system features, "raising of the front lax vowels" (e.g., *bet* → *bit*) and "the nonfronting of the high- and mid-back vowels" (e.g., not merging the vowels in LOT and THOUGHT) (Farrington, King & Kohn, 2021, p. 1). Based on Thomas's (2001; 2007) research on the AAVS and their own follow-up work, Farrington, King, and Kohn (2021) argue that the AAVS likely developed due to migration (e.g., The Great Migration), segregation, and place and identity (p. 1). Given these unique, and highly racialized, circumstances under which the AAVS developed, it is possible that speakers in Black communities pronounce vowels in words differently than speakers in other communities (Farrington, King & Kohn, 2021; Thomas, 2001; 2007). Saying *pen* as [pɪn] rather than [pɛn], for example, is associated Black speech (as well as southern speech) (Charity, 2008).

While sociolinguists have tested whether speakers' use of AAVE features correlates with listeners' judgments of speech as "sounding Black" (Gaither et al., 2015; Thomas & Reaser, 2004; Perrachione et al., 2010), they have not yet tested whether speakers' use of AAL features (i.e., articulation rate, front-stressing on initial syllables, and vowel pronunciation) correlates with listeners' judgments of speech and speakers as "sounding Black." Thus, I systematically analyzed and compared the 200 samples of Black and Non-Black speech along all three measures to address this gap.

6.4 Method

To determine whether the AAL features distinguish Black- from Non-Black-sounding speech, I used time aligned text grids and automated Praat scripts to extract duration, intensity, and formant values from the 200 PBJ4 and 1-20 speech samples. Then, I conducted statistical analyses in Microsoft Excel and *R* with the extracted values. The process I used to create the time aligned Praat text grids as well as the Praat scripts are available in Appendices C through H.

6.4.1 Articulation Rate

For the articulation rate analysis of the 200 speech samples, I used an automated Praat script to extract the number of phonemes, duration of silence, and total duration of each audio file. The data were read to an Excel file where I subtracted the duration of silence from the total duration of each speech sample to calculate the duration of speaking portions only. Then I divided the duration of these speaking portions by the number of phonemes in each clip to calculate the articulation rates (in phonemes/sec). I ran multiple linear regression analyses in R for each set of speech samples (PBJ4 and 1-20) to assess whether speakers' articulation rates (key independent variable) predicted their Black racialization scores (dependent variable). I ran the first models with speakers' race as a control variable and in the second models, I also included age, gender, socioeconomic status, and U.S. division as additional controls.

6.4.2 Front-Stressing of Initial Syllables

I analyzed six words in the 1-20 speech samples to determine whether front-stressing on initial syllables predicted speakers' Black racialization scores (calculated in Chapter 4). I chose to use only the 1-20 speech samples for this analysis in order to compare speakers' stress patterns in the same words and control for the effect that the word itself may have on the way in which speakers stress syllables. I selected all two-syllable *teen* words for the analysis (i.e., thirteen, fourteen, fifteen, sixteen, eighteen, nineteen) because these words' standardized American English pronunciations have primary stress on the second syllable, *-teen* (Oxford English Dictionary, 2023). Therefore, speakers who put stress on the first syllable, that is, speakers who front-stress, will be distinguishable from those who do not.

I used syllable duration and intensity as measures of stress based on research showing that stressed syllables in words are "longer and stronger" than unstressed syllables (e.g., Foulkes & Docherty, 2006, p. 409; Kelso et al., 1985). Therefore, stressed syllables have a longer duration and greater intensity than their unstressed counterparts. To extract the duration (ms) and mean intensity values (db) in the syllables, first, I modified each of text grids by hand to add a syllable tier. I developed automated Praat scripts to extract the duration and intensity of each syllable in the six key words. I calculated the duration ratio of each word by dividing the first syllable's duration by the second syllable's duration–and I did the same for intensity. Thus, for both duration and intensity, ratios greater than one indicated that speakers stressed the first syllable more than the second (i.e., front-stressing) and ratios less than one indicated that speakers stressed the second syllable more than the first (i.e., not front-stressing) (See section 6.5.2). I ran a series of regressions to test whether speakers' duration and intensity ratios (key independent variables) predicted speakers' Black racialization scores (dependent variable). For the first models, I included the key

independent variable (i.e., duration ratios or intensity ratios) and speakers' race (on a Black-Non-Black binary) as a control variable. And for the second models, I used the same key independent variables as well as the following controls: race (Black/Non-Black) age, gender, socioeconomic status, and U.S. division.

6.4.3 Vowel Pronunciation

I analyzed speakers' vowel pronunciations by comparing speakers' vowel formant arrangements. Formants are concentrated acoustic energy around a frequency in a speech wave (Ladefoged, 2006). On a spectrogram, formants look like dark bands of energy that expand across the duration of a vowel. The F1 band reflects vowel height—the higher the band, the lower the vowel's height. The F2 band reflects the degree of backness—the higher the band, the more fronted the vowel; and the F3 band reflects roundness of the vowel—the lower the band, the rounder the vowel (Ladefoged, 2006). Taken together, the arrangements of these formants relative to each other indicates which vowel is being produced. For example, [ɪ] and [ɛ] are both front, unrounded vowels; therefore, each vowel has similarly located F2 and F3 values. However, [ɛ] is slightly lower than [ɪ], so [ɛ]'s F1 band is relatively higher than /ɪ/'s. This is modeled below in the spectrograms of these two vowels (Ladefoged, 2006).

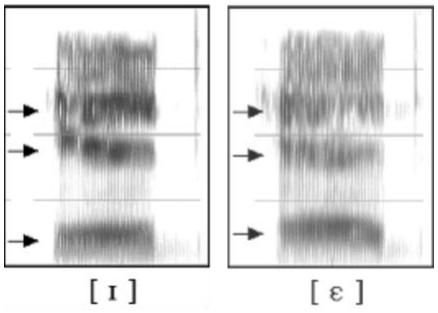

Figure 5 Spectrograms of [ɪ] and [ɛ] from Ladefoged (2006 p. 185)

I analyzed speakers' pronunciation of vowels by comparing normalized formant arrangements in speakers' vowels in each unique syllable in the words *one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen,* and *twenty*.[10] I chose to analyze the vowels in these syllables to control for the effect of phonetic environment on vowel production. In other words, because a vowel's phonetic environment can change its formant patterns (Hillenbrand & Clark, 2001), I chose to analyze speakers' vowels only when they were in the same phonetic environment, or a unique syllable. By "unique syllable," I mean that I did not individually test all 32 syllables in the words one through twenty. Rather, I combined identical syllables across the words into one category. For example, I considered the /ɛ/ in *SE.ven* and /ɛ/ in *SE.ven.teen* to be one category because the same vowel is in the same phonetic environment (i.e., the same unique syllable) in both words. Similarly, I collapsed /i/ in *thir.TEEN*, *four.TEEN*, *fif.TEEN*, *six.TEEN*, *seven.TEEN*, *eigh.TEEN*, and *nine.TEEN* into one category. However, I did not collapse the /i/ in *THREE* with the /i/ in the "teen" numbers because they are located in different phonetic environments. Taken together, I compared vowels in 19 unique syllables. These are listed in Table 8.

Traditional vowel analysis methods involve capturing F1, F2, and F3 at each vowel's midpoint, normalizing these values, calculating the differences between these normalized values (i.e., normalized F2-F1 and F3-F1), and plotting the vowels using these differences as x and y values (Thomas & Kendall, 2015). However, because these methods use formant values captured at one point in the vowel's duration, they do not take into account changes in how the vowel is

[10] Details on the normalization procedure I used are explained in the following three pages.

produced across its duration. For example, the mid-point approach cannot distinguish a stable vowel from a diphthong.

Table 8 Unique Syllables in Numbers 1-20

Vowel	Phonetic Environment	Number
/ʌ/	w_n	*one*
/u/	t_	*two*
/i/	θr_	*three*
/ɔ/	f_r	*four, four(teen)*
/aɪ/	f_v	*five*
/ɪ/	s_ks	*six, six(teen)*
/ɛ/	s_	*se(ven), se(venteen)*
/ə/	v_n	*(se)ven,* *(ele)ven (se)ven(teen)*
/eɪ/	#_t	*eight, eigh(teen)*
/aɪ/	n_n	*nine, nine(teen)*
/ɛ/	t_n	*ten*
/ɪ/	#_	*e(leven)*
/ɛ/	l_	*(e)le(ven)*
/ɛ/	tw_lv	*twelve*
/ɚ/	θ_	*thir(teen)*
/i/	t_n	*(thir)teen, (four)teen, fif(teen), (six)teen ,(seven)teen, (eigh)teen, (nine)teen*
/ɪ/	f_f	*fif(teen)*
/ɛ/	tw_n	*twen(ty)*
/i/	t_	*(twen)ty*

To take this into account, I extracted the formant values at the start and five percent intervals across the entire duration of each vowel rather than at the midpoint only. Therefore, I extracted 21 F1 values, 21 F2 values, and 21 F3 values for each of the 19 vowels for each speaker (N=119,700 extracted formant values in total). These values were automatically read to a .csv file using a Praat script.

The formant values I extracted were then normalized using *R* (Kendall & Thomas, 2010). Vowel normalization takes into consideration that speakers have different mouth sizes, therefore,

their formant resonances also differ (Disner, 1980; Thomas, 2002). Normalization, then, allows linguists to compare formant resonances in meaningful ways across speakers with different mouth sizes. According to Disner (1980) and Thomas (2002), in addition to reducing resonance variation caused by varying mouth sizes, normalizing vowels for analysis allows linguists to "preserve sociolinguistic/dialect/cross-linguistic differences in vowel quality…preserve phonological distinctions among vowels…[and] model the cognitive processes that allow human listeners to normalize vowels uttered by different speakers" (Thomas & Kendall, 2007). Thus, normalizing vowels is a necessary step to be able to compare formant frequencies across speakers while minimizing effects from speakers' unique physiological characteristics.

While there are several widely used vowel normalization methods, I used the Bark Difference Method (Syrdal & Gopal, 1986). Like all normalization methods, the Bark Difference Method (BDM) filters out physiological differences across speakers while maintaining sociolinguistic differences (Thomas & Kendall, 2007). However, and most importantly, the BDM is a vowel-intrinsic normalization method, meaning that all vowels in a speakers' repertoire are not needed for the method to work (Syrdal & Gopal, 1986). Because the vowels in the numbers one through twenty do not include all speakers' vowels, it was necessary to use the BDM.

To analyze the vowels using the BDM, first, the F1, F2, and F3 values for each vowel were extracted at five percent intervals and converted to Bark using the Traunmüller (1997) formula: $Z_i = 26.81/(1+1960/F_i) - 0.53$ in the vowels R package (Kendall & Thomas, 2010), where F_i is the value for a given formant, *i*, at a specified point in the vowel's duration and Z_i is the normalized value for F_i. These values were plotted using Z_2-Z_1 and Z_3-Z_1 as *y*-values and the time points (i.e., five percent intervals) as corresponding *x*-values. Best-fit lines were drawn through the points using *R* and the slopes and intercepts of each line were extracted for comparison using multiple

linear regression with and without controls. The first regression models included the slopes and intercepts as key independent variables along with race as a control variable to predict speakers' Black racialization scores (dependent/response variable). The second models also included age, gender, socioeconomic status, and U.S. division as control variables.

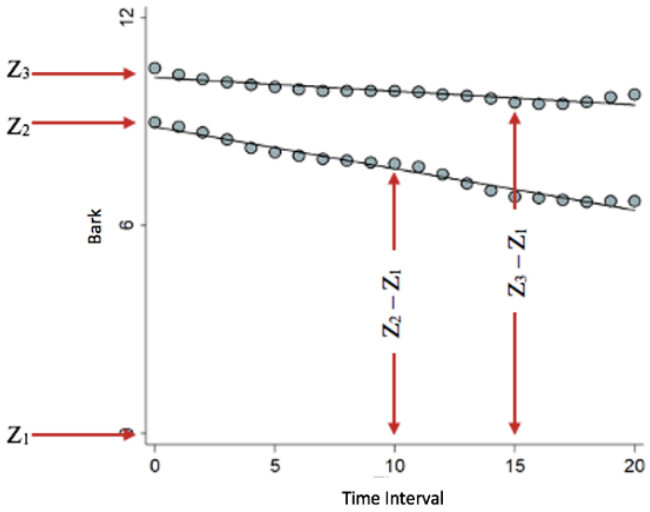

Figure 6 Normalized Vowel Plot for Black Female Speaker Saying "Ten"

Using the methods outlined above, I tested whether speakers' articulation rates, front-stressing on initial syllables, and vowel pronunciations predicted speakers' Black racialization scores.

6.5 Results

6.5.1 Articulation Rate

I conducted a series of multiple linear regression tests on both the PBJ4 and 1-20 files to determine whether speakers' articulation rates predicted speakers' Black racialization scores. The

first models included the key independent variable (PBJ4 articulation rate and 1-20 articulation rate) and speakers' race (Black or Non-Black) and the second models included the key independent variable and speakers' race, age, gender, socioeconomic status, and U.S. division as controls. All four regressions showed that speakers' articulation rates did not predict their Black racialization scores. In other words, this suggests that speaking more slowly did not trigger listeners to categorize a speaker as "sounding Black." The models used here, and elsewhere throughout this chapter, include the same control variables as in Chapter 5. The test results show that, like in Chapter 5, speakers' self-identified race and age remain significant predictors of listeners' voice racialization of speakers—that is, speakers' articulation rates did not affect the effect of speakers' race and age on listeners' voice racialization.

Table 9 Coefficients for Variables Predicting Black Racialization Scores - Articulation Rate (PBJ4)

	Model 1	*Model 2*
Articulation Rate	0.01 (0.01)	0.01 (0.01)
Speaker Characteristics		
Black	0.42 (0.05)*	0.39 (0.05)*
Age		0.00 (0.00)*
Gender (Ref. Man)		
Transgender Man		-0.22 (0.22)
Woman		0.03 (0.05)
Socioeconomic Status		-0.03 (0.03)
U.S. Division (Ref. New England)		
Middle Atlantic (NY, NJ, PA)		0.01 (0.12)
Eastern Midwest		-0.05 (0.11)
Western Midwest		-0.02 (0.12)
Eastern South		0.05 (0.10)
Deep South (KY, TN, MS, AL)		0.08 (0.14)
Western South (TX, OK, AR, LA)		0.22 (0.12)
Mountain		0.07 (0.14)
Pacific (WA, OR, CA, HI, AK)		-0.02 (0.13)
Constant	-0.05 (0.12)	-0.26 (0.22)
R^2	0.4785	0.5867
Number of Clips Categorized	100	

*Denotes that the coefficient is statistically significantly different from zero with a p-value <0.05

Table 10 Coefficients for Variables Predicting Black Racialization Scores - Articulation Rate (1-20)

	Model 1	*Model 2*
Articulation Rate	0.01 (0.01)	0.00 (0.01)
Speaker Characteristics		
Black	0.42 (0.05)*	0.24 (0.05)*
Age		0.00 (0.00)*
Gender (Ref. Man)		
Transgender Man		0.25 (0.23)
Woman		0.01 (0.05)
Socioeconomic Status		-0.02 (0.03)
U.S. Division (Ref. New England)		
Middle Atlantic (NY, NJ, PA)		0.10 (0.12)
Eastern Midwest		0.05 (0.12)
Western Midwest		0.05 (0.13)
Eastern South		0.09 (0.11)
Deep South (KY, TN, MS, AL)		0.06 (0.14)
Western South (TX, OK, AR, LA)		0.22 (0.12)
Mountain		0.06 (0.15)
Pacific (WA, OR, CA, HI, AK)		-0.05 (0.13)
Constant	0.15 (0.12)	-0.10 (0.23)
R^2	0.4806	0.5571
Number of Clips Categorized	100	

*Denotes that the coefficient is statistically significantly different from zero with a p-value <0.05

6.5.2 Front-Stressing of Initial Syllables

Next, I conducted a series of regression tests on the duration and intensity ratios of the first and second syllables in the words *thirteen*, *fourteen*, *fifteen*, *sixteen*, *eighteen*, and *nineteen* to assess whether speakers' syllable stress patterns predicted their Black racialization scores. The speakers' average duration and intensity ratios for each of the six words are provided in Table 11 below. The speakers' average duration ratios indicate that, in general, the *teen* syllable is longer than the first syllable (with the exception of *sixteen*). This means that, when considering duration alone, speakers are, for the most part, not front-stressing. However, when it comes to intensity of

the syllables, the first syllables in *thirteen*, *fourteen*, *eighteen*, and *nineteen* have greater intensities than the second syllables—suggesting that the words *are* front-stressed. On the other hand, *fifteen* and *sixteen* have greater intensities in the second syllables compared to the first syllables—suggesting that the words are not front-stressed. Taking the average duration and intensity ratios together, with the exception of *sixteen*, the speakers' duration and intensity ratios do not tell a consistent story about the way in which speakers stress syllables in these numbers.

Table 11 Mean Duration and Intensity Ratios

Word	**Duration Ratio (M*ean, SD*)**	**Intensity Ratio (*Mean, SD*)**
Thirteen	0.54, 0.20	1.10, 0.09
Fourteen	0.64, 0.20	1.08, 0.11
Fifteen	0.75, 0.26	0.97, 0.09
Sixteen	1.21, 0.40	0.97, 0.08
Eighteen	0.43, 0.17	1.12, 0.11
Nineteen	0.78, 0.20	1.08, 0.10

Like the articulation rate analysis, the first regression models in front-stressing analysis included the key independent variable (syllable duration or intensity ratios) and speakers' race (Black or Non-Black) and the second models included the key independent variable and speakers' race, age, gender, socioeconomic status, and U.S. division. All tests but one showed that the syllable duration and intensity ratios did not predict speakers' Black racialization scores. However, the intensity ratio of the syllables in *thirteen* was a statistically significant predictor of speakers' Black racialization scores ($p = 0.0364$) in the first model. In this model, larger intensity ratios for *thirteen*–which imply *front-stressing*–predicted higher Black racialization scores. When adding in the control variables, though, the intensity ratio for *thirteen* was no longer significant, but race and age were. This change suggests that, front-stressing *thirteen* does not actually predict speakers' Black racialization scores. Instead, it implies that speakers stress syllables in *thirteen* differently across age and race groups but controlling for these variables eliminates the effect of front-

stressing on speakers' Black racialization scores. Plainly, front-stressing *thirteen* does not predict speakers' Black racialization scores when race and age are taken into account. The regression results for *thirteen* are in Table 12 and the remaining test results are available in Appendix I.

Table 12 Coefficients for Variables Predicting Black Racialization Scores - *Thirteen* Intensity

	Model 1	*Model 2*
"Thirteen" Syllable Intensity Ratio	0.49 (0.23)*	-0.40 (0.23)
Speaker Characteristics		
Black	0.42 (0.04)*	0.39 (0.04)*
Age		0.00 (0.00)*
Gender (Ref. Man)		
Transgender Man		-0.08 (0.20)
Woman		0.02 (0.04)
Socioeconomic Status		-0.05 (0.03)
U.S. Division (Ref. New England)		
Middle Atlantic (NY, NJ, PA)		0.05 (0.10)
Eastern Midwest		0.00 (0.10)
Western Midwest		0.06 (0.11)
Eastern South		0.07 (0.09)
Deep South (KY, TN, MS, AL)		0.07 (0.12)
Western South (TX, OK, AR, LA)		0.20 (0.11)
Mountain		0.06 (0.13)
Pacific (WA, OR, CA, HI, AK)		-0.07 (0.12)
Constant	-0.46 (0.26)	-0.45 (0.30)
R^2	0.5288	0.5701
Number of Clips Categorized	**100**	

6.5.3 Vowel Pronunciation

The final series of regression tests included the slopes and *y*-intercepts of the Bark normalized F2 and F3 best-fit lines to determine whether speakers' pronunciation of vowels in each of the unique syllables in the numbers from one to 20 predicted speakers' Black racialization scores. Statistically significant results suggest that speakers' pronunciations differed and predicted listeners' voice racializations. In line with the previous analyses, the first regression models

included the key independent variable and speakers' race (Black or Non-Black), and the second models included the key independent variable and speakers' race, age, gender, socioeconomic status, and U.S. division. Of the 152 tests, one vowel showed statistically significant results across all four F2 measures tested with and without controls–the vowel in *TEN*. Normalized formant arrangements that aligned with an [ɪ]-like pronunciation predicted higher Black racialization scores whereas formant arrangements that aligned with an [ɛ]-like pronunciation predicted lower Black racialization scores—in alignment with earlier literature on race and the *pen-pin* merger (Charity, 2008). In other words, pronouncing *TEN* as [tɪn] predicted higher Black racialization scores. And pronouncing *TEN* as [tɛn] predicted lower Black racialization scores. The coefficients from the *TEN* regression tests are provided in Table (13) below and the remaining 144 test results are available in Appendix J.

Table 13 Bark Normalized Coefficients for Variables Predicting Black Racialization Scores - *TEN* Vowel

	Z2		Z3	
	Base Model	With Controls	Base Model	With Controls
y-intercept (start)	0.05*	0.04*	0.04*	0.02
Slope (change)	-0.82*	-0.62*	-0.44	-0.12

*Denotes that the coefficient is statistically significantly different from zero with a p-value <0.05

6.6 Discussion

The results of the analyses demonstrate that, across most of the measures, speakers' production of AAL features does not predict speakers' Black racialization scores—or listeners' voice racialization of the speakers. Simply put, speaking more slowly, front-stressing initial syllables, and AAVS vowel pronunciations did not, for the most part, explain whether a listener thought that a speaker "sounded Black." The results do, however, evidence the existence of a "pen-

pin merger" (i.e., /ɛ/ → [ɪ] / _N) among Black speakers in this study, which previous research has shown is heavily associated with speech in Black American communities (Charity, 2008; Green, 2002). In fact, in this sample, Black speakers are, in general, merging pen and pin more than Non-Black speakers. Accordingly, whether or not a listener thinks a speaker "sounds Black" is (significantly) predicted by speakers' use of [tɪn] or [tɛn] for *TEN*, with speakers using [tɪn] having higher Black racialization scores than speakers using [tɛn]. In other words, the results of this analysis suggest that listeners may have relied on speakers' pronunciation of *TEN* as [tɪn] to infer that the speaker they are hearing is Black.

Yet, the logic that listeners relied on speakers' pronunciation of *TEN* as [tɪn] to identify Black speakers is complicated by the fact that the *pen-pin* merger is also associated with speech in southern White communities (Austen, 2020; Bailey, 1997). Thus, the *pen-pin* merger is characteristic of southern speech more generally, across Black and White racial lines. In Chapter 4, my analyses demonstrated that listeners were significantly better at accurately identifying speakers' racial identities when the speakers were from southern divisions compared to other U.S. divisions, ($p < 0.05$). But in the south, pronouncing *TEN* as [tɪn] is associated with both Black and White speech. Thus, if listeners were relying on the *pen-pin* merger to identify Black speakers, they would have been *worse*, not better, at accurately distinguishing southern Blacks from Whites. In other words, it's more likely they would have misidentified White speakers as "sounding Black." Yet, the fact that listeners were *better* at distinguishing southern speakers' racial identities suggests that they did not rely on speakers' pronunciations of *TEN* to infer speakers' racial identities.

Taken together, the results of the analyses in this chapter as well as Chapter 5 support the argument that speakers' production of speech is not the only—and perhaps not the primary—factor

shaping how listeners perceive, categorize, and assess speakers/speech, even in cases where listeners' input is limited to speech alone. While it is certainly possible (in fact, probable) that I have not included AAVE and AAL features that would have significantly predicted listeners' assessments that speakers "sounded Black," I have included several key features that linguists have, for decades, shown are associated with speech in Black communities. Other prominent features, like AAVE/AAL morpho-syntactic features and lexical items were not prevalent in the dataset and, therefore, could not be tested. Moreover, since these variables were not prevalent in the dataset, they could not have influenced whether or not listeners categorized speakers as "sounding Black." In short, while other production features that I did not analyze may have shaped listeners' voice racialization of speakers, key AAVE/AAL features that have been previously documented in the literature as characterizing "Black-sounding" speech did not predict whether speakers "sounded Black" to the listeners in this study. In addition, other documented AAVE/AAL features were not prevalent in the dataset and, therefore, could not have predicted listeners' voice racialization of speakers nor could they have been tested.

In the context of this study's broader purpose to investigate whether "Black-sounding" speech is associated with criminality in the United States, the findings of this chapter, along with the findings in Chapter 5, indicate that speakers' production of AAVE or AAL features is unlikely to explain such an association. Therefore, in line with Inoue (2006), Flores & Rosa (2015), and Rosa & Flores (2017), moving forward in the rest of this my dissertation, I will no longer focus on the speech practices of racialized speakers.

6.7 Conclusion

The goal of this chapter was to demonstrate further that Black speakers' production of speech is likely not the primary factor shaping listeners' perception of speakers and their speech, even in cases in which they do not have additional information about speakers' identities. Specifically, in this chapter, I diverged from earlier research studies on the relationship between Blackness, language, and criminality by focusing on Black speakers' production of AAL features rather than those of AAVE, which mostly represent the speech practices of young Black men in the inner city (King, 2020; Smitherman, 1988; Morgan, 1994). Thus, I analyzed features that are associated with speech in Black communities more generally, rather than one subgroup of Black speakers. In alignment with my hypothesis, though, the majority of the features I analyzed did not correspond with listeners' voice racializations (i.e., speakers' Black racialization scores). And the one feature that was found to statistically significantly predict listeners' voice racializations–the pronunciation of the vowel in *TEN*–I argue that it is likely not a cue listeners are utilizing to determine whether a speaker is Black. Building off of the data and findings from this chapter as well as Chapters 4 and 5, in Chapter 7, I evaluate the extent to which listeners' perceptions of speakers' race—that is, speakers' Black racialization scores—correlate with listeners' perceptions of speakers' criminality.

7.0 Hearing Black and Remembering Crime: Listeners' Association of Black Racialized Speech with Criminality

7.1 Introduction

The United States of America is often referred to as the *land of opportunity* (Chiang, 2017), and many Americans believe that with hard work and perseverance any person, regardless of their race, can acquire success (Athreya & Romera, 2015; Chiang, 2017; Chetty et al., 2014; Hertel & Pfeffer, 2016). In fact, since the election of the nation's first Black president in 2008, many have championed the idea that America has totally abandoned its racist past and moved on to a "post-racial" society (Bonilla-Silva, 2015; Jost & Ogletree, 2011; Lum, 2009; Tesler, 2020; Tesler & Sears, 2010)–one where every person is judged based on the content of their character rather than the color of their skin (Bonilla-Silva, 2017; Carson & Shepherd, 2001). Post-racial ideology is bolstered by the fact that, since the 1950s, Americans have expressed fewer explicitly racist views than their counterparts in decades past (Bonilla-Silva, 2021; Dovidio & Gaertner, 1991, Moberg, Krysan & Christianson, 2019; Schuman et al., 1997; Sheatsley, 1966). Yet, the daily experiences of racially minoritized and marginalized U.S. residents demonstrate that race is still an important factor determining social status and outcomes in America (Alexander, 2020; Bonilla-Silva, 2021; Howell & Korver-Glenn, 2018; Kozol & Perluss, 1992; Oliver & Shapiro, 2013; Massey & Denton, 1993; Shapiro, 2004). Put another way, despite the fact that Americans explicitly express fewer racist views than before, racism continues to be a significant issue in the United States. But what explains this inconsistency?

Scholars have illuminated that "invisible" forms of racism (Braveman et al., 2022, p. 171), like institutionalized racism and implicit racial associations, explain how racism persists in the United States even while most Americans denounce it (Braveman et al., 2022; Bonilla-Silva, 2017; Brownstein & Zalta, 2019). The present study is an experimental investigation of one example of this–Americans' persistent, but invisible, association of Blackness with criminality. While racial attitudes are multidimensional and complex, surveys have demonstrated that the prevalence of anti-Black views among White Americans specifically has generally declined since the collapse of Jim Crow (Peffley & Hurwitz, 1998; Moberg, Krysan & Christianson, 2019). Despite this, the over-representation of Black Americans at every point in the U.S. criminal legal system–from profiling to arrests and guilty verdicts to sentencing and post-incarceration discrimination–evidences that anti–Black racism still exists in American society (Alexander, 2020; Hinton & Cook, 2021; Johnson & Lee, 2013; Kovera, 2019; Pager, 2003; Sommers & Marotta, 2014; Yang, 2015). Even so, the disparities in the criminal legal system are just one manifestation of Black criminalization in the United States. In fact, they are part and parcel of a 300-year-old anti-Black punitive tradition that criminalizes Black Americans throughout many institutions and U.S. society, not just in courts and prisons (Hinton & Cook, 2021)

While this anti-Black punitive tradition manifests itself in different ways across time and space, sociolinguistic research has demonstrated that listeners' association of Black speech (i.e., AAVE) with criminalized factors (e.g., guilt, lack of credibility, untrustworthiness) shapes Black speakers' experiences and social outcomes (Corley [Branson], 2014; Jones et al., 2019; Kurinec & Weaver III, 2019; Rickford & King, 2016). For example, a 2019 study found that court reporters' implicit associations of AAVE speakers with "criminality, deviance, and untrustworthiness" (Jones et al., 2019, p. e247) was reflected in their transcripts of Black speakers' testimonies,

potentially impacting verdict decisions. This study and others like it highlight that listeners' associations may contribute to the criminalization of Black AAVE speakers in the courtroom (Corley [Branson], 2014; Jones et al., 2019; Kurinec & Weaver III, 2019; Rickford & King, 2016).

In the preceding chapters, I highlighted the importance of shifting from speakers' production of AAVE (or other language varieties/features associated with Black Americans) to a perception-based heuristic, *Black racialized speech,* and demonstrated that listeners' perception of speakers does not always line up with how speakers physically produce speech. In this chapter, I evaluate the extent to which listeners' perception of speakers' race correlates with their perception of speakers' criminality. However, I build from earlier studies by examining whether and to what extent this association exists *beyond* the courtroom and in broader society,[11] and I also gauge whether this association exists despite Americans' belief in post-racial ideology. This contextualization to broader society would provide valuable insight to the ways in which listeners' perception contributes to practices that criminalize Black Americans and help to develop effective strategies to mitigate and dismantle these practices.

For this study, I developed and implemented an online memorization experiment to test whether 100 U.S.-based listeners (henceforth "participants") associated images of crime relevant objects with Black racialized speech. For the experiment, each participant listened to a Black racialized speech sample and then completed a memorization task. The memorization task involved participants viewing a screen with 18 crime relevant (e.g., gun, handcuffs) and crime irrelevant objects (e.g., heart, sunglasses) and then recounting all the objects they remembered in 30 seconds. I predicted that participants who heard speech samples with higher Black racialization

[11] The importance of doing this was emphasized earlier, in Chapter 2.

scores would remember more crime relevant objects than participants who heard speech samples with lower Black racialization scores. Overall, higher Black racialization scores did not correlate with the number of criminalized objects participants remembered ($r = 0.021$). However, I found a small positive correlation between these variables among participants who heard Black speakers' speech samples ($r = 0.213$). In other words, higher Black racialized speech scores correlated with remembering more criminalized objects, but only when the speaker was Black–and this was despite the fact that participants were not informed of speakers' racial identities. In addition, the majority of participants expressed belief in post-racial ideology, suggesting that the participants unfairly associate Black racialized speech with criminality, even while they believe that America is fair for all.

This chapter proceeds as follows: first, I define and outline the differences between overt and covert (i.e., "invisible") forms of racism, including implicit associations. Second, I lay out research from psychology and psycholinguistics, which shaped the experiment I designed to test whether listeners associate Black racialized speech with criminality. Third, I describe the experimental study, including the study's purpose, material's design, participant recruitment, survey deployment, data collection, and analytical methods. Finally, I present the results of the study, discuss their implications, and conclude.

7.2 Background

Today, the majority of Americans do not express explicit anti-Black views (Anti-Defamation League, 1993; Bonilla-Silva, 2017; Dovidio & Gaertner, 1991, Moberg, Krysan & Christianson, 2019; Schuman et al., 1997; Sheatsley, 1966). Yet anti-Black racism continues to be

an important issue in the United States (Alexander, 2020; Bonilla-Silva, 2017; Howell & Korver-Glenn, 2018; Kozol & Perluss, 1992; Oliver & Shapiro, 1995; Massey & Denton, 1993; Shapiro, 2004). Many scholars explain this inconsistency by distinguishing overt (i.e., "old fashioned") racism from covert (i.e., "invisible") racism, which reinforces persistent racial inequalities in less obvious ways than outright racism (Braveman et al., 2022; Bonilla-Silva, 2017; Brownstein & Zalta, 2019). In this section, I disentangle "old fashioned" racism (Bobo & Fox, 2003) from "invisible" racism (Braveman et al., 2022) (e.g., institutionalized racism, implicit associations), while delineating how each contributes to structures, processes, and practices that reinforce the association of Blackness with criminality in the United States.

7.2.1 'Old Fashioned' Racism

Overt racism, or what Bobo and Fox (2003) and others (e.g., Virtanen & Huddy, 1998) call "old fashioned" racism, is characterized by outright racial prejudice and beliefs that racial groups should be segregated in society–for example, that individuals from different racial groups should not attend the same schools or marry. It is typified by the ideology that not only do alleged biological differences (e.g., genetics) distinguish racial groups, but that these differences, and thus racial groupings, are hierarchically ranked (Smedley & Smedley, 2005). Much of the research on old fashioned racism in the United States has focused on Whites' perceptions of Blacks and the view that Black Americans are innately inferior to White Americans. Some of the most prominent early work on this centered around Blacks' purported genetic propensity toward criminality (e.g., Lombroso, 1876). For example, Lombroso (1876) outlined the alleged physical characteristics of "The Criminal Man"–a biological class of humans who were supposedly innately criminal–which he said resembled "the Negroid" (p. 245). Simply put, he contended that Africans and their

descendants were born criminals and his work was influential in popularizing the pervasive construct of the 'Black criminal' in the United States shortly after the abolition of slavery.

In 1910, however, Du Bois countered Lombroso's claims by underscoring the ways in which society's definitions of race and crime contributed to perceptions of Black Americans as criminals. He explained how vagrancy laws, which were put into place immediately following Emancipation, criminalized recently freed Black Americans for being unemployed and relegated them back to an exploitative slave-like system through the Black Codes and convict leasing (Du Bois, 1910; Blackmon, 2009). Thus, Du Bois's work demonstrated that race is a social construct and not a biological fact, which is how most social scientists view race today (Morning, 2014; Smedley & Smedley, 2005). And like the majority of social scientists, most Americans no longer explicitly express old-fashioned views that racial groups are innately hierarchically ranked based on genetics (Bonilla-Silva, 2021; Kinder & Sanders, 1996). However, scholars contend that this ideological shift did not eliminate racism in America (Bobo & Fox, 2003; Bonilla-Silva, 2021)—and this is reflected by the reappearance of 'old fashioned' forms of racism in America since the election of Donald Trump in 2016, which has diminished (but not totally eradicated) the overall "positive trend in White's racial attitudes" since Jim Crow (Flores-González & Salgado, 2021, p. 342). Nevertheless, "invisible" racism–like institutionalized racism and implicit associations–characterizes racism in the United States today (Braveman et al., 2022).

7.2.2 Invisible Racism

Covert racism, or what Braveman et al. (2022) call "invisible" racism, is perpetuated or enacted unconsciously, implicitly, or in ways that are not easily visible (Braveman et al., 2022). To clarify, this is not to say that the *impact* of invisible racism is difficult to see; rather, that it

operates in less direct ways than outright racism. Braveman et al. (2022) describe invisible racism as "systemic and structural…[and] pervasively and deeply embedded in and throughout systems, laws, written or unwritten policies, entrenched practices, and established beliefs and attitudes that produce, condone, and perpetuate widespread unfair treatment of people of color" (p. 171). Based on this definition, invisible racism can be delimited into two types: institutionalized (e.g., systems, laws, policies, practices) and personally mediated (e.g., beliefs, attitudes) (Jones, 2000).[12]

Institutionalized racism operates via policies and practices embedded within organizations and institutions (e.g., hospitals, schools, courts) that unfairly disadvantage racially minoritized groups while benefiting White Americans (Jones, 2000). And personally mediated racism operates via associations that rationalize unfair treatment of racially minoritized people, including implicit associations[13] (i.e., unconscious bias) (Bobo & Fox, 2003; Jones, 2000; Payne, Vuletich & Brown-Iannuzzi, 2019). While distinguishing invisible racism into these two types helpfully illustrates different ways racism operates in U.S. society today, they are not mutually exclusive. In fact, the beliefs and attitudes that characterize personally mediated racism are by-products of institutionalized inequalities of the past and present (Jones, 2000; Payne, Vuletich & Brown-Iannuzzi, 2019). Thus, personally mediated racism must be contextualized within a broader schema that incorporates institutionalized racism rather than framed as solely individual biases. As such, below, I outline institutionalized racism first and then explain personally mediated racism, while focusing on implicit racial associations.

---

[12] Other scholars have also identified *interna*lized racism–"internalization of racial oppression by the racially subordinated" (Pyke, 2010, p. 551).

[13] Scholars have identified other forms of personally mediated racism, such as symbolic racism, laissez faire racism, and principled politics (Bobo & Fox, 2003).

7.2.2.1 Institutionalized Racism

Institutionalized racism, also called structural or systemic racism, refers to policies and practices embedded within organizations and institutions, like hospitals, schools, and courts, that disadvantage racially minoritized groups while benefiting White Americans (Pyke, 2010). Institutionalized racism is not perpetuated by individuals based on their attitudes or beliefs, which is why it is not always easy to identify in practice, but it can be identified by its profound impact and reinforcement of racial disparities through the United States.

When it comes to Black criminalization, simply overviewing racial disparities in arrests, incarceration rates, and sentencing illuminates institutionalized racism throughout the criminal legal system. For example, at age 18, Black men's likelihood of being arrested is 30 percent while White men's likelihood is 22 percent. At age 23, the likelihood of arrest for Black men escalates to 49 percent, while for White men the likelihood of arrest increases to only 38 percent (Schleiden et al., 2020). Moreover, five times more Black Americans than Whites are incarcerated; twice as many Black women than White women are imprisoned; one in every three Black men is issued a felony conviction in his lifetime; and while Black youth constitute only 14 percent of U.S. children, they nevertheless represent 32 percent of the youth arrested, 42 percent of the youth detained, and 52 percent of the youth tried in America's criminal rather than juvenile courts (NAACP, 2021). Finally, convicted Black Americans' sentences are, on average, ten percent longer than their White counterparts (Rehavi & Starr, 2014). Taken together, these disparities clearly show the ways in which institutionalized racism is embedded within the U.S. criminal legal system.

7.2.2.2 Personally Mediated Racism

Personally mediated racism, on the other hand, refers to the enactment of prejudice and discrimination by individuals. Jones (2000) defines prejudice as, "differential assumptions about

the abilities, motives, and intentions of others according to their race" and discrimination as, "differential actions toward others according to their race" (p. 1212-1213). Thus, discrimination is one manifestation of prejudice. Personally mediated racism can occur consciously, like old fashioned racism[14], or unconsciously, like implicit racial associations (i.e., unconscious bias). However, as emphasized earlier, implicit racial associations are a reflection of past and present forms of institutionalized racism and not solely individuals' biases (Jones, 2000; Payne, Vuletich & Brown-Iannuzzi, 2019).

7.2.2.3 Implicit Racial Associations (Unconscious Bias)

Implicit associations, also called unconscious bias, refers to individuals' unconscious connections between concepts (i.e., racial groups) and stereotypes or judgments that reinforce social hierarchies (e.g., race, gender, socioeconomic status) (Brownstein & Zalta, 2019). For example, implicit associations between Blackness and criminality help to explain why White Americans are fearful of Black Americans, even when they are simply engaging in harmless everyday activities (Thurston, 2019). These associations are acquired through individuals' social environment. Therefore, they do not always reflect individuals' endorsement of the associations or biases; rather, they reflect the associations these individuals have been exposed to (Karpinski & Hilton, 2001). Thus, explaining the inconsistency between Americans' declining anti-Black views, belief in post-racial ideology and the persistence of anti-Black racism in society.

Implicit associations are difficult to identify because they are unconscious—people make connections without realizing it. But psychologists, including psycholinguists, have innovated

[14] Scholars have outlined other conscious forms of racism besides old fashioned (i.e., overt) racism. For example, symbolic racism, laissez faire racism, and principled politics (Bobo & Fox, 2003) which operate using an alleged racial neutral mediator.

ways to expose these connections and unveil the ways in which individuals' associations contribute to persistent racial inequalities (e.g., Eberhardt et al., 2004; Eberhardt et al., 2006). These innovative methods influenced the experiment I designed to test whether U.S. residents associate Black racialized speech with criminality. Below, I overview the two studies that significantly contributed to my experiment design.

7.2.3 Implicit Association Experiments in Psychology and Psycholinguistics

7.2.3.1 Contributions from Psychological Research

Psychologists have innovated methods to create experiments and test individuals' implicit racial associations. I drew from Eberhardt et al.'s 2004 study, *Seeing Black*, to create an experiment to test whether U.S. listeners associate Black racialized speech with criminality without specifying the association to the courtroom. In their study, the research team conducted an experiment to assess whether participants associated Black faces with crime relevant objects. Specifically, they wanted to know whether *priming* participants with Black faces decreased their perceptual threshold for identifying images of crime relevant objects. Priming research participants prior to having them complete an experiment or task allows researchers to gain insight to participants' associations without their conscious awareness. Therefore, the experiment results can illuminate implicit associations that participants are not, themselves, aware of (Bermeitinger, 2016; Molden, 2014).

Eberhardt and her team randomly assigned participants to a prime type (i.e., Black, White, or no-prime control) and object type (i.e., crime relevant or crime irrelevant). For the Black and White primes, they used 50 images of Black and White male faces while controlling for height, weight, age, and attractiveness. Participants in the Black and White prime conditions were primed

with 110 face images, each appearing as a "flash" on the screen for 30ms. For the object types, they used black and white line drawings of knives and guns for the crime relevant objects, and everyday household items (e.g., a book, camera, and cup) for the crime irrelevant objects. The line drawings of the objects were transformed into short videos with 41 frames. The first frame of each video was a digitally degraded image of a crime relevant or irrelevant object which was edited to look pixelated, fuzzy, and unidentifiable. Each subsequent frame was increasingly clearer such that the final frame was the original unedited line drawing (Eberhardt et al., 2004).

Despite the fact that participants were unaware they were seeing Black or White faces, participants who saw Black faces were able to identify the crime relevant objects sooner than participants who saw White faces. And participants' explicit racial attitudes had no effect on this finding (Eberhardt et al., 2004). Simply put, priming the participants with Black faces made it easier for them to identify crime relevant objects, suggesting they associate Black faces with crime or criminality.

I drew out two key elements from Eberhardt et al.'s study, which I implemented into my experiment design–priming race and testing participants' association between the prime and images of crime relevant objects as a measure of criminality. First, using the priming technique helped to illuminate participants' implicit associations. Similarly, I chose to prime participants with samples of Black racialized speech (i.e., the extent to which a speakers' speech "sounds Black") rather than Black and White faces. Second, by priming with crime relevant objects (e.g., gun, knife) the study was situated to examine the relationship between Blackness and criminality more broadly, and not just in the criminal legal system. Thus, to examine the relationship between Black racialized speech and criminality, I decided to use images of crime relevant objects in my study instead of conducting a court-related experiment. Due to conducting my experiment

remotely instead of in-person, I did not follow Eberhardt et al.'s methodology exactly. The areas where I diverged from the team's methods are detailed and explained in Section 7.3 (Methods) below.

7.2.3.2 Contributions from Psycholinguistics

Priming research participants audibly rather than visually, as in Eberhardt et al.'s (2004) study, requires a slightly different approach. Therefore, I needed additional input regarding how to implement speech samples into the experiment to implicitly prime racial associations. I drew from Gather et al.'s 2015 study, *Sounding Black or White: Priming Identify in Biracial Speech*, to gain insight on this. In the study, Gaither et al. (2015) investigated whether biracial speakers' speech practices aligned with one or the other of their racial identities after being primed with a Black- or White-sounding speech sample. They predicted that the biracial speakers in their study would alter their speech such that it aligned with the racial identity they were primed to hear. The results of the study showed that the biracial speakers *did* alter their speech in ways that aligned with the racial identity they were primed to hear. In addition, the researchers found that audio recordings of 3-4 minutes as well as 10-20 seconds in length were both sufficient to prime associated racial identities (Gaither et al., 2015). Finally, they used a set of coders who were not biracial to determine whether listeners' own racial identities influenced the ways they perceived Black- and White-sounding speech. Yet, like the findings I outlined in Chapter 4, listeners' own racial identities did not influence the ways they perceived Black- and White-sounding speech (Gaither et al., 2015). In the context of my study, Gaither et al.'s work shows that audio recordings are viable for use as primes to elicit implicit racial associations. Moreover, having listeners hear 10-20 seconds of Black- or non-Black-sounding audio is sufficient to prime the associated racial identities among listeners.

Taking together these contributions from Eberhardt et al. (2004) and Gaither et al. (2015), I designed a memorization experiment to test whether U.S. listeners associate Black racialized speech with criminality as described just below.

7.3 Method

I designed an online memorization experiment using Qualtrics, and I distributed the Qualtrics survey to a convenience sample of 100 participants on Prolific. In what follows, I give an overview of the experiment design and analytical methods for the present study.

7.3.1 Survey Design – Qualtrics Survey

I tested listeners' implicit associations using a memorization experiment I designed in Qualtrics. I distributed the Qualtrics survey (seven sections) to a convenience sample of 100 Prolific participants. The first section was a Welcome page with a summary of the experiment and its purpose as well as a question for participants to indicate their consent to participate. If a participant indicated that they did not want to participate in the study, they were redirected to the end of the survey and then back to Prolific. However, participants who consented to the study were moved forward to the CAPTCHA verification page. This verification helps to prevent non-human subjects from attempting to complete the survey. On this page, participants simply checked a box confirming that they are not a robot. Next, participants provided their Prolific IDs. However, I set up the Qualtrics survey to automatically capture participants' Prolific IDs from their URLs. Yet, if there was a problem with this, participants could still input their Prolific IDs manually. By

including this step in the survey, I was able to connect the participants' Qualtrics IDs to their Prolific IDs, linking each participant's demographic information in Prolific to their responses to the Qualtrics survey.

After participants' Prolific IDs were captured, they moved on to a page with the study's first set of instructions. On this page, they saw the following message: *On the next page, you will hear an audio clip. Listen to the entire clip. Click the arrow when you are ready to hear the clip.* After clicking the arrow, participants heard one of the 1-20 audio clips[15] automatically, ensuring that they heard the entire clip. After listening to the audio clip, participants moved forward to the memorization task. For this task, participants saw a screen with 18 objects—five crime relevant objects and 13 crime irrelevant objects—and were later asked to recall all of the objects they remembered from the screen. The purpose of this task was to determine whether participants were more likely to recall criminalized objects when they were primed with higher-scoring Black racialized speech samples compared to lower-scoring Black racialized speech samples. If participants did remember more criminalized objects when primed with Blacker-sounding speech, this would suggest that they may associate Black racialized speech with criminality.

On the first page of the task, participants saw instructions stating, *On the next page you will see 18 images. Memorize as many as you can. You will have 10 seconds. Click the arrow when you are ready to begin.* After clicking the arrow, they saw the screen with the 18 objects (see figure 7 below) for 10 seconds. The set of images was created using publicly available icons in Microsoft PowerPoint. Of the objects displayed, I categorized the police officer, gun, handcuffs, gavel, and dice as 'crime relevant' and all others I considered 'crime irrelevant.'

[15] I chose to use the 1-20 clips, rather than the PBJ4 clips, based on the experiment design from Gaither et al. (2015), as these were all at least 10 seconds in length.

Figure 7 Image of Crime Relevant and Crime Irrelevant Images

After viewing the image for 10 seconds, participants were automatically moved forward to another page where they were given the following instructions: *On the next page, list all items you remember (one item per box). You will have 30 seconds. Click the arrow when you are ready to begin.* After clicking the arrow, they saw a page with 18 boxes, which appeared for 30 seconds. Here, participants input the title of each item they remembered in a separate box. After 30 seconds, the memorization task was completed and participants moved on to the demographics questions which asked for participants' gender, racial identity, ethnicity, religious affiliation, education level, household income, current zip code, and childhood county of residence. The last question was a measure of participants' belief in post-racial ideology. The question asked participants to indicate the extent to which they agreed with the statement: *In America, success is determined by how hard you work, not your identity (for example, your gender, ethnicity, family history, etc.).*[16] To answer this question, they could choose from the options: *Strongly Disagree, Somewhat Disagree, Neither*

[16] Despite important differences between race and ethnicity, I chose to use "ethnicity" rather than "race" for this question because it is generally thought to be a less sensitive word than "race," and therefore, less likely to influence participants' responses.

Agree nor Disagree, *Somewhat Agree*, and *Strongly Agree*. After responding to this question, participants saw a screen where they were thanked for their participation and then rerouted back to Prolific.

7.3.2 Analytical Method

For the analysis, I set out to determine whether listeners who were primed with higher-scoring Black racialized speech samples were more likely to remember crime relevant objects compared to participants who were primed with lower-scoring Black racialized speech samples. When I say this, I simply mean that I tested whether there was a positive relationship between speakers' Black racialization scores and the proportion of crime relevant objects participants remembered. In other words, I wanted to know whether listeners who heard speakers that were *more likely* to be categorized as "sounding Black" (i.e., have a relatively higher Black racialization score) remembered more crime relevant objects than their counterparts who heard speakers that were *less likely* to be categorized as "sounding Black" (i.e., have a relatively lower Black racialization score). To determine this, I collected participants' responses to the memorization experiment, weighted each object participants remembered based on the object's frequency in the dataset, used the weights to calculate the proportion of criminalized objects each participant remembered, and conducted a Pearson's *r* test to determine whether higher-scoring Black racialized speech samples correlated with higher proportions of criminalized objects remembered. Prior to doing these calculations, though, I cleaned and coded the data to account for the fact that participants used different words to refer to the same image.

To clean and code the data, I used Excel's UNIQUE(tocol) function to create a list of all unique words in the dataset. Reviewing this list, I found that several participants listed items that

did not represent one of the 18 objects. For example, a few participants listed “dog” as one of the objects they remembered, but an image of a dog was not included. All of these misremembered words were removed from the dataset. After removing the misremembered words, I categorized the remaining words into 18 categories based on the 18 objects and replaced the individual words in the dataset with their corresponding category title using Excel. These categories were: PIZZA, DICE, DINOSAUR, BASKETBALL, HANDCUFFS, PLANE, BARN, GUN, CAT, FISH, GAVEL, SCISSORS, CAR, HEART, BONE, ACORN, SUNGLASSES, and COP, and the words I included in each category are provided in Table 14.

Table 14 Word Categorization for Memorization Experiment

Category	Words Included in Category
PIZZA	piazza, pizza, pizza slice, slice o
DICE	dice
DINOSAUR	brontosaurus, dinosaur, dinosar, dinosour
BASKETBALL	basketball, ball, basjetbalk, baskey, bask
HANDCUFFS	cuffs, handcuff, handcuffs, hand cuffs
PLANE	airplane, jet plane, plane
BARN	barn, farm, home, house
GUN	gun, handgun
CAT	cat
FISH	fish, goldfish
GAVEL	gavel, hammer, mallet
SCISSORS	scissor, scissors, sissors, sissocers
CAR	car, csr
HEART	heart
BONE	bone, dog bone
ACORN	acorn
SUNGLASSES	sunglasses, glasses
COP	polic, police, police man, policeman, police officer, cop

After replacing all words in each category with their corresponding category titles, I highlighted all of the criminalized categories (i.e., DICE, HANDCUFFS, GUN, GAVEL, and COP) in green to easily identify these items later on in the analysis.

Next, to account for the fact that some objects were more memorable than others, I weighted each image based on its frequency in the dataset. To do this, I used the COUNTIF function in Excel to count the number of times each of the category titles appeared in the dataset. Then, I divided each of these numbers by the total number of category titles (for all categories) in the dataset to determine the proportion of times each category appeared in the dataset. To give greater weight to less frequent items, I subtracted each of these proportions from 1—reversing the weights. Then, I replaced each category title with its corresponding weight.

The weights were used to score each participants' memory of crime relevant objects. This score, which I call the *criminalization score*, was calculated for each participant. I did this by summing the weights of crime relevant objects a participant remembered and dividing this by the sum of the weights of all objects the participant remembered. In other words, the participant's criminalization score is the weighted proportion of criminalized objects the participant remembered (compared to the sum of the weights of all objects the participant remembered).

$$\text{Criminalization Score} = \frac{\text{Weighted Sum of Crime Relevant Objects Remembered}}{\text{Weighted Sum of All Objects Remembered}}$$

Since each participant listened to one, and only one, speech sample, each participant's criminalization score was paired with the corresponding speaker's Black racialization score. And I conducted Pearson's *r* tests on the full dataset, a subset with only Black speakers, and subset with only Non-Black speakers to determine whether higher-scoring Black racialization scores correlated with higher criminalized image scores. If so, this would suggest that listeners implicitly associated Black racialized speech with crime relevant objects. I also conducted three regression tests to predict participants' criminalization scores using the speech samples' Black racialization scores at the key independent variable and the post-racial ideology measure as a control. I

operationalized the post-racial ideology by converting participants' responses to scores from 1-5 (*strongly disagree* = 1, *strongly agree* = 5).

7.4 Results

I distributed the experiment to a convenience sample of 100 Prolific participants on March 18, 2023 at 8:00am EST and received all responses at 9:30am EST. While I distributed my survey to only 100 participants, an error in Prolific resulted in one extra participant taking the study—for a total of 101 participants. Due to using a convenience, rather than representative, sample, the participants' demographics did not reflect the U.S. population more broadly. However, since all of the research participants are U.S. residents this provides some insight to whether or not, U.S. residents generally associate Black racialized speech with criminality. The participants' demographics are summarized in Table 15.

Table 15 Participants' Demographics

	Mean	Standard Deviation	Min	Max
Participant Characteristics				
Racial Identification				
White	0.71	0.46	0	1
Black	0.05	0.22	0	1
Latinx	0.02	0.14	0	1
Asian	0.12	0.33	0	1
Indigenous	0.01	0.10	0	1
Multiracial	0.09	0.30	0	1
Age	39.10	14.17	18	72
Gender				
Man	0.49	0.50	0	1
Non-Binary	0.02	0.14	0	1
Woman	0.50	0.50	0	1
Socioeconomic Status	3.23	0.83	1	5
Number of Participants		**101**		

7.4.1 Image Weighting and Scoring

Together, the 101 participants in the study recalled 644 objects—about 6.4 objects per participant—not including the words they misremembered (e.g., dog). I organized the 644 words into 18 categories using the categorization scheme I outlined earlier. Then, I calculated the frequency of each category in the dataset. COP was remembered most frequently—65 times. While SUNGLASSES was remembered least frequently—only 13 times. Therefore, COP had the smallest weight while SUNGLASSES had the greatest weight. The categories, along with their frequencies and corresponding weights are provided in Table 16.

Table 16 Object Categories, Frequencies, and Weights

Category	Frequency	Weight
HEART	57	0.9114906832
ACORN	23	0.9642857143
PIZZA	29	0.9549689441
BASKETBALL	39	0.9394409938
SCISSORS	19	0.9704968944
FISH	16	0.9751552795
SUNGLASSES	13	0.9798136646
BARN	43	0.9332298137
DINOSAUR	23	0.9642857143
PLANE	45	0.9301242236
CAT	57	0.9114906832
COP	65	0.899068323
BONE	22	0.9658385093
DICE	34	0.9472049689
CAR	33	0.948757764
GUN	51	0.9208074534
HANDCUFFS	55	0.9145962733
GAVEL	20	0.9689440994

Using these weights, I calculated the criminalization scores for the participants. The average criminalized image score for the 101 participants was 0.3841. However, the lowest criminalized image score was 0 and the highest was 0.7443. I paired participants' criminalized image scores with the corresponding speakers' Black racialization scores and conducted a Pearson's *r* correlation test on the full dataset. The results did not show a correlation between participants' criminalized image scores and speakers' Black racialization scores ($r = 0.0206$). I conducted two more Pearson's *r* tests—one test that included only Black speakers and another that included only Non-Black speakers. For Black speakers, the test showed a positive correlation between participants' criminalized image scores and speakers' Black racialization scores ($r = 0.2132$), while, for non-Black speakers, the test did not show a correlation ($r = -0.0186$).

I conducted three multiple linear regression tests to determine whether the Pearson *r* coefficients were statistically significant and to assess whether participants' degree of agreeance with post-racial ideology (*mean* = 3.35, *SD* = 1.01) corresponded with the weighted proportion of crime relevant objects they remembered. The first regression model predicted participants' criminalized image scores using the speech samples' Black racialization scores (key independent variable) and participants' responses to the post-racial ideology question (control variable). Neither variable statistically significantly predicted the criminalized image scores. The second model predicted participants criminalized image scores using the same variables but for the subset of Black speakers only. Again, neither independent variable statistically significantly predicted the criminalized image scores. Finally, the third model was set up the same way as the previous two but included the subset of only Non-Black speakers. The variable incorporating participants'

responses to the post-racial society question, however, was a statistically significant predictor of participants' criminalization scores ($p < 0.05$).

Table 17 Coefficients from Multiple Linear Regressions Predicting Prop. of Crime Objects Remembered

	All Speakers	Black Speakers	Non-Black Speakers
		Mean (St. Err.)	
Black Racialization Score	0.34 (0.17)	0.11 (0.07)	0.03 (0.18)
Post-Racial Society Measure	0.00 (0.01)	-0.04 (0.02)	0.03 (0.02)*
Constant	0.37 (0.05)*	0.43 (0.08)*	0.29 (0.06)*
R^2	-0.0196	0.0576	0.0500
Number of Participants		**101**	

* Denotes that the coefficient is statistically significantly different from zero with a p-value < 0.05

7.5 Summary and Discussion

Thus far in the dissertation, I have built a corpus of spoken language from Black and Non-Black U.S. residents, captured listeners' racialization of speakers (and speech), and operationalized these judgments to construct the Black racialized speech heuristic. I did this to importantly shift the focus from production to perception so that, in this chapter, I could properly investigate the role of linguistic practices—that is, *listeners' perception*—in U.S. Black criminalization. While earlier sociolinguistic research on this has focused on speakers' use of AAVE in courtroom contexts, I wanted to know whether and to what extent listeners' practices contributed to Black criminalization in broader society. To determine this, I drew from psychological and psycholinguistic research studies to design a memorization experiment and determine whether a convenience sample of 100 U.S. residents associated Black racialized speech with images of crime relevant objects. Overall, the study did not find that participants associated Blacker-sounding voices with criminality. However, when testing Black and non-Black speakers separately, results

showed a small, positive correlation between higher-scoring Black racialized speech samples and higher criminalization scores—but only for Black speakers. In fact, for Black speakers, even though the regression coefficient is not statistically significant, the model predicted that listeners would remember 10.60 percent more criminalized items for the highest scoring Black racialized speaker compared to the lowest scoring Black racialized speaker. This suggests that when listeners hear Black speakers who are also voice-racialized as 'sounding Black', they may associate what they hear with criminality. Simply put, this sample of U.S. residents may associate Blacker-sounding speech with criminality when they are listening to a Black speaker—even when they *do not know* the speaker is Black and despite the fact that most participants believe America is fair for all.

7.6 Conclusion

The goal of this chapter was to determine whether U.S. listeners associate Black racialized speech with criminality without delimiting the context to courtrooms. I tested this by conducting an experiment using the 1-20 speech samples collected previously (see Chapter 3) and found a small, but insignificant correlation between these variables among participants who heard Black speakers' samples. While it is not statistically significant, the group of listeners who heard Black speakers did remember more crime relevant objects when primed with higher-scoring Black racialized speech samples. In the Final Discussion and Conclusion chapter, which follows, I highlight the social implications of this work, especially contextualizing the findings of this study with the findings from Chapter 5 and 6, which showed that several key AAVE and AAL features

that I tested did not predict speakers' Black racialization scores. I also discuss next steps for using this research as a springboard for initiating change in research practice and beyond.

8.0 …*Let Them Hear:* Final Discussion and Conclusion

8.1 Introduction

The purpose of this dissertation was to address the question, *Do U.S. listeners associate Black racialized speech with criminality?* While an emerging body of sociolinguistic research has already demonstrated that linguistic practices are indeed implicated in Black criminalization processes, the work focused on speakers' vocal production of AAVE and has centralized courtroom contexts (Corley [Branson], 2014; Kurinec & Weaver III, 2019; Jones et al., 2019; Rickford & King, 2016). This framing does not take into account the ways in which listeners' perception shapes what they 'hear' and thus, how they evaluate Black speakers (Flores & Rosa, 2015; Rosa & Flores, 2017), nor does it consider the fact that Black criminalization is pervasive in U.S. society and manifests in spaces outside of the criminal legal system (Hinton & Cook, 2021). To address these gaps, I developed and distributed three online surveys, conducted hundreds of statistical analyses, and laid out the findings in six empirical chapters to draw one crucial conclusion: The findings suggest that U.S. listeners may associate Black racialized speech with criminality, even when speakers are not using some prominent AAVE and AAL features documented in the literature (see Chapters 5 and 6). However, it appears that listeners may make this association only when they are listening to Black speakers—despite not knowing speakers' self-identified race or having other, external cues to decipher their racial identities.

In this final chapter of the dissertation, I review the findings in each of the previous empirical chapters and synthesize them to demonstrate how I drew the previously stated conclusion. Then, I discuss limitations and questions that remain unanswered, while highlighting

the contributions of my dissertation to the body of research on the relationship between Blackness, language, and criminality in the United States. Afterwards, I pan out to overview this study's relevance to the field of sociolinguistics as well as broader society. And finally, I conclude with strategies to move forward towards racial justice using linguistics (Charity Hudley, Mallinson & Bucholtz, 2020).

8.2 Summary

To address my overarching research question, I began by contextualizing the present study within the broader history Black criminalization in the United States (Chapter 2). Doing this allowed me to illuminate the ways in which Blackness and criminality have been and continue to be mutually co-constructed throughout all of American society (Hinton & Cook, 2021). This contextualization exposed a gap in earlier sociolinguistic literature, which focused on the relationship between Blackness, language, and criminality in the courtroom only (Corley [Branson], 2014; Kurinec & Weaver III, 2019; Jones et al., 2019; Rickford & King, 2016). And while there are many benefits to studying language and race in this context, the historical review of Black criminalization showed that this is only one space where Blackness is criminalized. In turn, I decided not to delimit the present study to the courtroom.

After historicizing the study, I explained how I built a corpus of naturalistic speech samples from Black and Non-Black speakers across the United States (Chapter 3). Having naturalistic speech samples from a wide range of speakers to implement in the proceeding studies increased the ecological validity of the work by more accurately reflecting language in the real world (Campbell-Kibler, 2010). Once I had the speech samples, I implemented them into an online

survey to collect listeners' racialized assessments of the speakers–a process I call, *voice racialization* (Chapter 4). I operationalized these judgments to construct the Black racialized speech variable, which I chose to use instead of AAVE or AAL. I decided to do this based on research in raciolinguistics and psycholinguistics that emphasizes the importance of listeners' perception, over speakers' vocal production of speech, in social categorization processes (Alim et al., 2016; Drager, 2010; Flores & Rosa, 2015; Hay & Drager, 2010; Levon, 2014; Niedzielski, 1999; Rosa & Flores, 2017; Rubin, 1992). I supported this decision by conducting over 100 statistical analyses across 38 AAVE (Chapter 5) and AAL (Chapter 6) production-based variables, which, for the most part, did not statistically significantly predict whether listeners thought speakers "sounded Black."[17] I did find, however, that one variable statistically significantly predicted whether a listener voice-racialized a speaker as "sounding Black"—the vowel in the number *TEN* (Chapter 6). Yet, I also explained that it is unlikely listeners used this feature as a cue to determine speakers' racial identities, because listeners were best at distinguishing Black from Non-Black speakers in the south, where many speakers from different racial backgrounds pronounce *TEN* the same way (Bailey, 1997).

Finally, I deployed a memorization experiment to test whether U.S. listeners associated Black racialized speech with criminality in broader U.S. society (Chapter 7). I accomplished this by implementing images of five crime relevant objects into the memorization experiment in which I tested listeners' association between Black racialized speech and criminality (Chapter 7). Overall, I found that when listeners heard Black speakers, higher Black racialization scores correlated with higher criminalized image scores. Taking into consideration that speakers' speech production does

[17] It is important to note that, in general, speakers did not use many AAVE and AAL features. This may explain the lack of correlation. Additional research must be conducted to determine whether the lack of AAVE and AAL features prevalent in the data explains these findings.

not affect the extent to which listeners think speakers "sound Black" (i.e., their Black racialization scores) (Chapter 5 and Chapter 6), I attributed this association to listeners' own perception. Thus, I concluded that U.S. listeners may associate Black racialized speech with criminality, even when speakers are not using some prominent AAVE and AAL features documented in the literature (see Chapters 5 and 6). However, it appears that listeners may make this association only when they are listening to Black speakers—despite not knowing speakers' self-identified race or having other, external cues to decipher their racial identities.

8.3 Limitations

Despite the findings of my research study and their suggested implications, like all empirical research, my dissertation has limitations. I delineate the limitations of my dissertation into two broad groups: methodological and analytical. In what follows, I explain the nature of these limitations, justify the choices I made, and provide suggestions as to whether and how to address these gaps.

8.3.1 Methodological Limitations

A primary goal of my dissertation was to add to the current body of sociolinguistic research on the relationship between Blackness, language, and criminality and to contextualize the work by investigating how linguistic practices contribute to Black criminalization in broader society. I investigated this using the memorization experiment I outlined in Chapter 7. However, the sample for this experiment had only 101 participants and, therefore, is not large enough to make

generalizable statements about U.S. listeners more broadly. I chose to use a sample this size because memorization tasks have not been utilized in prior research on this topic; therefore, first, I needed to determine whether the method was effective in capturing the relationship between Black racialized speech and criminality before proceeding further. Having shown that there is a correlation between speakers' Black racialization scores and participants' criminalized images scores, there is evidence to suggest that the method is effective. At this juncture, the work should be scaled up, in consultation with other researchers with expertise in psychological methodology (e.g., psycholinguists), to determine whether the association between Black racialized speech and criminality is more widespread beyond the present sample of 101 participants.

One other goal of my dissertation was to collect samples of speech and judgments from participants across the United States. This required distributing my surveys remotely, with no interaction between me and the participants. Recently, however, sociolinguists have advocated for methods that actively engage participants in research and ensure that participants, especially Black Americans, receive tangible benefits from the work (above and beyond financial compensation and general policy implications). Therefore, an important next step would be complementing this research with qualitative, participant-led research out in the field. This could include in courtrooms, schools, other institutions, or in the public where researchers interact with speakers in broader society. Importantly, though, the research must be led by participants who, themselves, take ownership of the work and determine when, where, and how it moves forward.

8.3.2 Analytical Limitations

Another important goal of my dissertation was to empirically demonstrate that speakers' physical production of speech does not shape listeners' social categorization of them. Specifically,

I wanted to demonstrate that use of AAVE features does not predict speakers' Black racialization scores. Yet, I conducted a listening assessment to determine the extent to which speakers utilized six AAVE phonological features; therefore, my assessment could be affected by the same listening biases I illuminate in this study. Although a more objective approach, like taking formant, duration, or intensity measurements, for example, could less subjectively capture how speakers produce speech, I do not recommend this approach. Throughout this dissertation, I have emphasized the importance of listeners' perceptions in shaping their judgments of racialized speakers. Therefore, I do not endorse any approach that would continue to dissect the speech practices of racialized speakers to try to determine what listeners are attending to when they criminalize them. Like Morrison (1975), I argue that this is a distraction. Rather, researchers should devote energy to further investigating how listeners' listening practices are implicated in processes that racialize and criminalize Black Americans, as there has been far too little empirical research on listeners' perceptions (and the factors that influence these perceptions) not only in the present body of research on Blackness, language, and criminality, but in sociolinguistics more generally.

Although my dissertation has the above-mentioned limitations, what it contributes to the emerging body of work on the relationship between Blackness, language, and criminality, I argue, adds important value to the field of sociolinguistics and also has broader impacts for society with respect to advancing racial justice. The present study importantly highlights that Black speakers' criminalization is delimited neither to their use of AAVE nor to the courtroom. And listeners may associate Black racialized speech with criminality (for Black speakers) even when they do not know speakers' racial identities. These contributions could potentially shape the direction of research on this topic, especially as it pertains to further investigating the important role of listeners' listening practices in processes that racialize and criminalize Black Americans.

8.4 Contributions to Sociolinguistics

The present study has importantly highlighted the need for more sociolinguistic research focusing on the critical role of listeners' perception in social categorization processes, like racialization, as well as in enactments of discrimination, prejudice, and racism, like criminalization. In addition to highlighting this need, my dissertation demonstrated one strategy for operationalizing listeners' perceptual judgments to investigate how listening practices are implicated in these social processes. By capturing multiple listeners' judgments and operationalizing them using the method I outlined in Chapter 4, other sociolinguists can implement listener-centered linguistic measures into their research to illuminate the ways in which listening practices impact speakers' outcomes.

To identify the need to focus on listeners' perception in this work, however, I had to draw heavily from disciplines outside of sociolinguistics including other linguistic subdisciplines, like raciolinguistics and psycholinguistics, as well as other social science fields like sociology, psychology, critical race studies, and critical criminology. While sociolinguists, like many scholars, are siloed among themselves, my dissertation highlights the benefits of incorporating research from outside disciplines in sociolinguistic research. Moving forward, I urge sociolinguists to collaborate with researchers in other disciplines–especially as it pertains to research on race and racism–to continue to improve the potential impact of the work.

8.5 Contributions to Society

In addition to the above-mentioned contributions to sociolinguistics, the work also has implications for advancing racial justice efforts in broader society. As has been illustrated throughout this dissertation, Black criminalization is one of the most pervasive forms of anti-Black racism in America, impacting Black Americans' daily experiences and outcomes since colonial times and continuing to the present (Alexander, 2020; Hinton & Cook, 2021; Oshinsky, 1997). This dissertation has illuminated that using AAVE/AAL features may not be the only—or even the most important—factors shaping whether listeners think a speaker "sounds Black." Moreover, it suggests that listeners may associate Black racialized speech with criminality when they are hearing Black speakers—even if they do not know the racial identities of the speakers they are listening to. These findings help to counter the widely held belief that race is seen but not heard and therefore, racism cannot be enacted through hearing alone. The present study has demonstrated that not only do listeners' listening practices racialize Black speakers, but they may also be implicated in processes that also criminalize Black speakers. Thus, this work has illuminated that, to dismantle anti-Black racism in all its forms in America, listeners' listening practices must be called out, addressed, and reformed.

8.6 Lingering Questions

At the outset, I wanted to address the question of whether U.S. listeners associate Black racialized speech with criminality. I have demonstrated that when speakers self-identify as Black, listeners may associate their voices with criminality, even when the listeners do not know speakers'

own racial identities. However, additional questions remain. Specifically, how do the findings of this research translate to the real world? After all, the experiments in this study tested listeners' racialization and criminalization of speakers using disembodied voices and that is not how language typically works in society. First, though, I would like to acknowledge that there are many important scenarios in which Black criminalization occurs, even while listeners cannot see speakers physically. For example, 9-1-1 dispatchers do not see the callers who are seeking assistance, yet some dispatchers are required to assess the likelihood that the caller is responsible for the emergency or crime for which they are seeking help (Harpster, Adams & Jarvis, 2009). In other words, they are required to gauge the caller's criminality. In light of the conclusions drawn from this study, this could have devastating consequences for how an emergency is communicated to the police, how police arrive on the scene, if and how a person receives medical care, and whether an innocent person is arrested.

Nonetheless, in many everyday instances of Black criminalization, listeners have other contextual information from which to draw to contextualize speakers' racial identities and make judgments about their criminality. Investigating the ways in which broader context shapes listeners' assessments of Black speakers is beyond the scope of this dissertation. But it is an important question to address to more fully understand the relationship between Blackness, language, and criminality beyond empirical studies.

8.7 Future Directions and Conclusion

Future research directions should focus on addressing the role of listeners' listening practices in social categorization and criminalization processes. This can be accomplished by

continuing to conduct sociolinguistic research that investigates these practices–especially research going beyond the Black-Non-Black binary which was utilized in this study. Importantly, though, listening practices can be addressed through institutional change. As I acknowledged in Chapter 7, personally mediated racism, like implicit racial associations, is the result of past and present structural and institutionalized inequalities (Jones, 2000; Payne, Vuletich & Brown-Iannuzzi, 2019). Therefore, biased listening practices cannot be fully eradicated by focusing on changing individuals' attitudes. A more effective approach is to directly challenge policies and practices that contribute to racial inequality and implement policies that denaturalize and destabilize the association of Blackness (including Black racialized speech) and criminality. Sociolinguists can contribute to these efforts by importantly shifting from focusing on production to *perception*.

Appendix A Qualtrics, Phonic, and Prolific Integration Process

Below, I have outlined the steps I used to create the Qualtrics survey, including the process I used to integrate Qualtrics, Phonic, and Prolific. The first page of the Qualtrics survey was an informed consent page that briefly described the recording process and told speakers what types of devices they could use to create the recordings. Audio recordings made with personal devices vary in terms of their faithfulness to recordings made with professional equipment, dependent upon the type of device that is used (i.e., tablet, smartphone, or laptop) (Freeman & De Decker, 2021). For example, laptops and smartphones perform best. However, recordings made with tablets are not as faithful to recordings made with professional equipment. They can still be useable, though, depending on researchers' goals (Freeman & De Decker, 2021). Based on prior research on self-recordings, all types of devices were sufficient for my study's goals (Freeman & De Decker, 2021).

On the second page of the survey, I included a CAPTCHA verification. This verification asked speakers to check a box on the screen as an additional security layer that prevents bots from trying to complete the survey. On the third page, I integrated my Qualtrics survey with Prolific. To do this, I added a text entry question asking speakers for their Prolific IDs—strings of letters and numbers that are anonymously linked to the research participants on Prolific. I allowed speakers to enter their Prolific IDs manually, but I set up the question to automatically capture their IDs from the URL through a process called 'Embedded Data'. To set this up, first, I labeled the block with my text entry question 'Prolific ID'. Then, for the question text, I including the

following message suggested on the Prolific website: "What is your Prolific ID? *Please note that this response should auto-fill with the correct ID*" (Prolific Team, 2022).

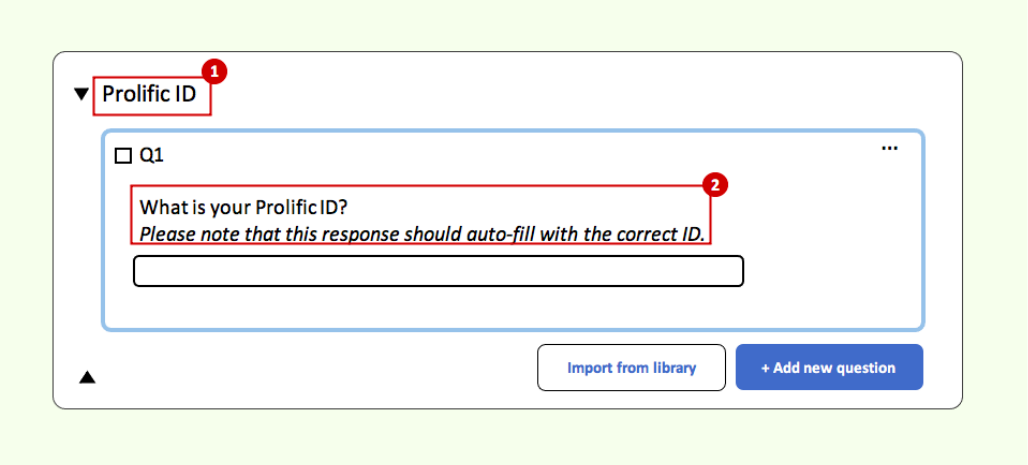

Figure 8. Prolific ID Question in Qualtrics Survey[18]

Next, I added an 'Embedded Data' element at the beginning of the survey in the 'Survey Flow' tab. Once the element was added, I entered 'PROLIFIC_PID' into the embedded data field and ensured that the field said, "Value will be set from panel or URL".

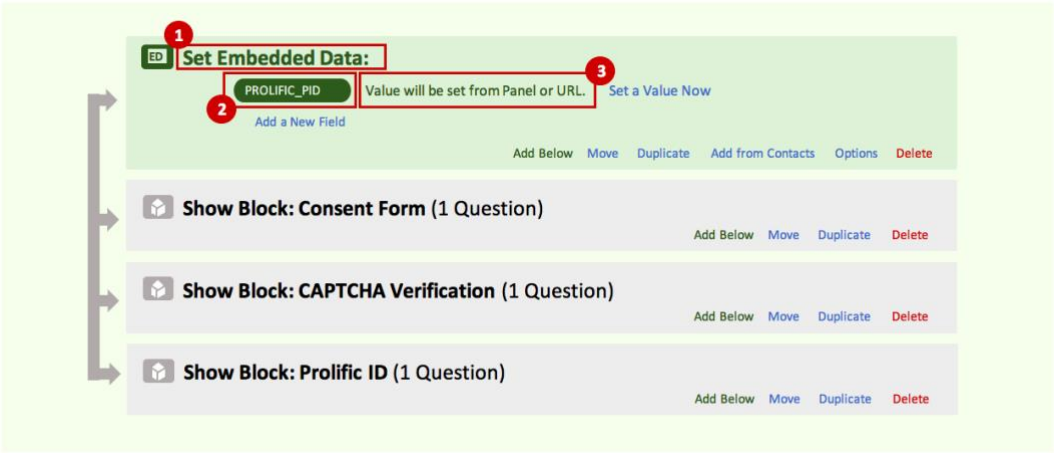

Figure 9. Adding an 'Embedded Data' Element in Survey Flow

[18] Instead of using screenshots, I recreated these images using Microsoft Word.

Then, I returned to the text entry box in the ‘Survey Builder’ tab, right clicked on the box, and chose ‘Default Choices’.

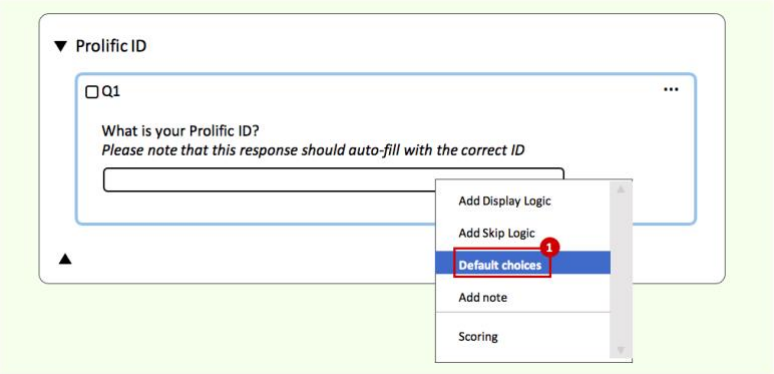

Figure 10. Selecting 'Default Choices'

On the window that popped up, I selected ‘Embedded Data’ and entered ‘PROLIFIC_PID’ in the search field on the drop box. After clicking ‘Insert’, the survey was set up to automatically capture speakers’ Prolific IDs and include these in the survey results.[19]

[19] A complete Qualtrics integration guide is available on the Prolific website. https://researcher-help.prolific.co/hc/en-gb/articles/360009224113.

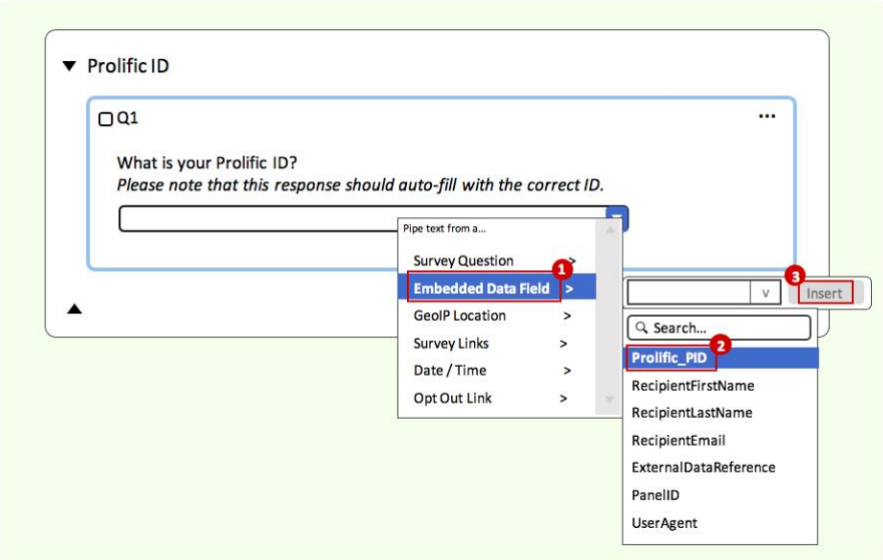

Figure 11. Converting Text Entry to 'Embedded Data Field' in Qualtrics Survey

By taking this step to collect speakers' Prolific IDs, I was able to connect the survey recordings to speakers' demographics. These included speakers' age, sex, racial identity, country of birth, country of residence, nationality, language(s), student status, and employment status. I also included additional demographic questions asking speakers for their gender, religious affiliations, income levels, current zip codes, and childhood state and county of residence.

On the fourth page of the survey, I included instructions on how to complete the audio recordings, which I asked speakers to complete in a quiet place with no background noise. I intended that my instructions were clear, easy to follow, and short. So, I used bold typeface to emphasize the most important steps and italics to supplement the bolded instructions. The following are my instructions exactly as I laid them out in the survey:

Instructions. The following four video clips show the steps to make a peanut butter and jelly sandwich. Watch a clip and then audio record yourself explaining it. If you have technical issues, push the HELP button for more instructions.

STEP 1: **Watch the clip.** *Press the white triangle on the video clip.*

STEP 2: **Start recording.** *Press the red "Record" button.*

STEP 3: **Record description.** *Speak naturally, as if you are talking to friends*.

STEP 4: **Stop recording.** *Press the red button again.*

STEP 5: **Replay recording (if desired).** *Press the small black triangle.*

STEP 6: **Submit.** *Click the submit button.*

On the fifth page of the survey, I added my first video prompt into the question text field. In this video a person was spreading peanut butter on a piece of bread. Then, I integrated a Phonic audio recording widget into the survey so that speakers could make their first recording. To do this, I had to sign into my Phonic account and create a skeleton survey. This skeleton survey included all five audio responses questions that I planned to include in my Qualtrics survey. After creating the skeleton survey, I published it by clicking the 'Publish' button in the top right corner on the website. This automatically advanced my browser to the main dashboard where I was able to extract the survey and question IDs to embed in my Qualtrics survey. I found these IDs by clicking on each question and locating the corresponding number strings in the URL of my browser. The first number string was the survey ID, which was the same for each question in the survey, and the number string at the end of the survey was the question ID—unique to each question. After locating these IDs in Phonic, I returned to my Qualtrics survey and clicked on 'Look and Feel' in the menu on the left side of my screen. Then I clicked 'General'. In the Header box, I pasted the Phonic code below, ensuring not to click 'edit' as this could have altered the code.

```
<script type="text/javascript" src="https://api.phonic.ai/phonic-
embed.min.js"></script>
```

Figure 12. Phonic Code Entered to Header Box

Moving forward, I returned to the 'Survey Builder' tab and located the question where I wanted to embed my first Phonic voice recorder into the survey. Then, I copied-and-pasted the

following Phonic code into the Qualtrics question's HTML editor (located under the question text), substituting "YOUR-SURVEY-ID" and "YOUR-QUESTION-ID" for the IDs extrapolated form the URL earlier.

```
<phonic-embed
  mode="audio"
  surveyid="YOUR-SURVEY-ID"
  questionid="YOUR-QUESTION-ID"
  responseid="${e://Field/ResponseID}"
  platform="qualtrics">
</phonic-embed>
```

Figure 13. Phonic Code to Integrate Recording Widget in Qualtrics

The recording widget was now successfully integrated to my survey and mandatory for speakers to complete before moving to the next question. I added in the recording widgets for my remaining four questions by repeating this integration process. Finally, I clicked 'Filters', then 'Display Options' and checked the box next to 'External ID' in Phonic to connect speakers' Qualtrics IDs to their recordings in Phonic, fully integrating all three online platforms.[20]

To further assist speakers with the recording process, I included a custom HELP button on each page with recording widgets. I did this by pasting the code below into the Qualtrics question text in while in HTML view, replacing "INSERT TEXT HERE" my instructions.

```
<button id="btn1"
style="position:absolute;left:50%;top:85%">HELP</button><span id="dialog"
title="Recording Instructions"> INSERT TEXT HERE</span>
```

Figure 14. HTML Code for Custom HELP Button

[20] The Phonic website has two integration guides for embedded recording widgets to Qualtrics surveys. https://docs.phonic.ai/docs/integrate-with-qualtrics and https://docs.phonic.ai/docs/how-do-i-customize-the-microphone-widget-settings

Appendix B Qualtrics Survey Design for Voice Racialization Study

The Qualtrics survey I used for the voice racialization study had five sections: Welcome, CAPTCHA Verification, Prolific ID, Voice Racialization Task, and Demographic Questions. On the welcome page, listeners were given a short description of the study and its purpose and were asked to consent to participating in the study. Those who did not consent to participate were redirected to the end of the survey; whereas those who did consent were sent to the CAPTCHA verification on the next page. The CAPTCHA verification page was included as a screening measure to ensure that only human participants completed the survey. On this page, listeners simply checked a box. The Prolific ID page automatically captured listeners' Prolific IDs from their browsers' URL, allowing me to link listeners' surveys to their demographic data in Prolific. I outlined the process of setting up the survey to automatically capture listeners' Prolific IDs from their URLs in Appendix A.

The Voice Racialization task consisted of 10 pages, each with one audio file (also "speech sample"). On each page, I set up Qualtrics randomization to randomly select one audio file from the final 200 audio files I used in this study (see Chapter 2 for more information on how I chose these files). I also added display logic to ensure that audio files used earlier in the survey could not be randomly selected later on. In other words, I set up the survey such that listeners heard and assessed (i.e., racialized) 10 different audio files and this set of 10 audio files was unique to each listener. Finally, the demographics section asked for listeners' age, racial identity, country of birth, country of residence, language(s), student status, employment status, gender, religious affiliations, income levels, current zip codes, and childhood state and county of residence.

To racialize the speech samples, on each page, listeners were asked to listen to the audio clip and then finish the sentence, *In my opinion, this voice sounds…* They were able to complete the sentence by choosing from the options *Asian, Black, Latinx, White,* or *A Different Group Not Listed*. If listeners chose '*A Different Group Not Listed*', they were given the additional choices *Hawaiian or Pacific Islander, Indigenous American or Alaskan, Middle Eastern or North African, Multiracial,* and *Another Group Not Listed (specify)* with a text entry box. I included another question wherein I captured listeners' confidence by asking, *How easy or difficult was it to choose?* They could choose one of the following options on a four-point Likert scale, *Very Difficult, Somewhat Difficult, Somewhat Easy,* and *Very Easy.*

Appendix C Time Aligned Text Grid Development Process

I used Phonic and Dartmouth's online Linguistic Automation (DARLA) tools to create time aligned Praat text grids using automated processes. Briefly, time aligned text grids can be created by uploading a verbatim transcript (.txt) and corresponding audio file to the DARLA website. To make the verbatim transcripts, I started with the automated transcripts provided by Phonic—one of the three online platforms I used to collect the speech samples—and then made corrections to the transcripts by hand while listening to the audio files. Taking the verbatim transcripts and corresponding audio files, I uploaded the pairs into DARLA one at a time to create time aligned text grids. Within minutes the files were sent to my email, where I downloaded them to my device and uploaded them in Praat. I viewed each transcript and corrected misaligned word boundaries by hand. I tested a sample text grids to determine the word boundary error rate. Out of 1,475 word boundaries, I found 52 misalignments. Overall, the automated text grid development process was effective and saved a significant amount of time.

Appendix D Praat Script for Total Duration of Sound File

```
writeInfoLine: "Getting duration of all syllables..."

# Specify the directory containing the sound and Textgrid files

directory$ = "/Users/dominiquebranson/Desktop/120/"

# Create the output file and write the first line

outputPath$ = "/Users/dominiquebranson/Desktop/120/Syllable_Duration.csv"
writeFileLine: "'outputPath$'", "file, syllable, duration"

# Make a list of all of the sound and Textgrid files in the directory and put the number of
filenames
# into the variable "number_files"

Create Strings as file list... list 'directory$'*.wav
number_files = Get number of strings

# Set up a "for" loop to iterate once for every file in the list

for j from 1 to number_files
select Strings list
current_token$ = Get string... 'j'
Read from file... 'directory$''current_token$'
```

```
# Make object_name$ variable that will equal the file name minus the extension

object_name$ = selected$ ("Sound")

# Read in file and extract duration for each non-empty tier

Read from file... 'directory$''object_name$'.TextGrid

select TextGrid 'object_name$'
number_of_intervals = Get number of intervals... 1
for b from 1 to number_of_intervals
        select TextGrid 'object_name$'
        interval_label$ = Get label of interval... 1 'b'

if interval_label$ <> ""
        begin_phoneme = Get starting point... 1 'b'
        end_phoneme = Get end point... 1 'b'
    duration = end_phoneme - begin_phoneme

# Save to spreadsheet
appendFileLine: "'outputPath$'",
                                ...object_name$, ",",
                                ...interval_label$, ",",
                                ...duration
endif
endfor
```

```
select all
minus Strings list
Remove
endfor

select all
Remove
clearinfo

    appendInfoLine: "Whoo-hoo! It didn't crash!"
```

Appendix E Praat Script to Extract Total Duration of Non-Empty Intervals

```
writeInfoLine: "Getting duration of all syllables..."

# Specify the directory containing the sound and Textgrid files

directory$ = "/Users/dominiquebranson/Desktop/120/"

# Create the output file and write the first line

outputPath$ = "/Users/dominiquebranson/Desktop/120/Syllable_Duration.csv"
writeFileLine: "'outputPath$'", "file, syllable, duration"

# Make a list of all of the sound and Textgrid files in the directory and put the number of
filenames
# into the variable "number_files"

Create Strings as file list... list 'directory$'*.wav
number_files = Get number of strings

# Set up a "for" loop to iterate once for every file in the list

for j from 1 to number_files
select Strings list
current_token$ = Get string... 'j'
Read from file... 'directory$''current_token$'
```

```
# Make object_name$ variable that will equal the file name minus the extension

object_name$ = selected$ ("Sound")

# Read in file and extract duration for each non-empty tier

Read from file... 'directory$''object_name$'.TextGrid

select TextGrid 'object_name$'
number_of_intervals = Get number of intervals... 1
for b from 1 to number_of_intervals
        select TextGrid 'object_name$'
        interval_label$ = Get label of interval... 1 'b'

if interval_label$ <> ""
        begin_phoneme = Get starting point... 1 'b'
        end_phoneme = Get end point... 1 'b'
    duration = end_phoneme - begin_phoneme

# Save to spreadsheet
appendFileLine: "'outputPath$'",
                                ...object_name$, ",",
                                ...interval_label$, ",",
                                ...duration
endif
endfor
```

```
select all
minus Strings list
Remove
endfor

select all
Remove
clearinfo

    appendInfoLine: "Whoo-hoo! It didn't crash!"
```

Appendix F Praat Script to Extract Syllable Duration Ratios

```
writeInfoLine: "Getting duration of all syllables..."

# Specify the directory containing the sound and Textgrid files

directory$ = "/Users/dominiquebranson/Desktop/Test/"

# Create the output file and write the first line

outputPath$ = "/Users/dominiquebranson/Desktop/Test/Syllable_Duration.csv"
writeFileLine: "'outputPath$'", "file, word, duration_ratio"

# Make a list of all of the sound and Textgrid files in the directory and put the number of
filenames
# into the variable "number_files"

Create Strings as file list... list 'directory$'*.wav
number_files = Get number of strings

# Set up a "for" loop to iterate once for every file in the list

for i from 1 to number_files
	select Strings list
	current_token$ = Get string... 'i'
	Read from file... 'directory$''current_token$'
```

```
        object_name$ = selected$ ("Sound")
        Read from file... 'directory$''object_name$'.TextGrid
        select TextGrid 'object_name$'
        nInt = Get number of intervals: 4
                for j from 1 to nInt
                        label$ = Get label of interval: 4, j

if label$ = "thirteen"
        start = Get starting point: 4, j
        end = Get end point: 4, j
        dur = end - start

syllone = Get interval at time: 3, start
syllone$ = Get label of interval: 3, syllone
startsyllone = Get starting point: 3, syllone
endsyllone = Get end point: 3, syllone
syllonedur = endsyllone - startsyllone

sylltwo = Get interval at time: 3, end - 0.001
sylltwo$ = Get label of interval: 3, sylltwo
startsylltwo = Get starting point: 3, sylltwo
endsylltwo = Get end point: 3, sylltwo
sylltwodur = endsylltwo - startsylltwo

durratio = syllonedur/sylltwodur

# Save to a spreadsheet
```

```
appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...durratio

endif

if label$ = "fourteen"
        start = Get starting point: 4, j
        end = Get end point: 4, j
        dur = end - start

syllone = Get interval at time: 3, start
syllone$ = Get label of interval: 3, syllone
startsyllone = Get starting point: 3, syllone
endsyllone = Get end point: 3, syllone
syllonedur = endsyllone - startsyllone

sylltwo = Get interval at time: 3, end - 0.001
sylltwo$ = Get label of interval: 3, sylltwo
startsylltwo = Get starting point: 3, sylltwo
endsylltwo = Get end point: 3, sylltwo
sylltwodur = endsylltwo - startsylltwo

durratio = syllonedur/sylltwodur
```

```
# Save to a spreadsheet

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...durratio

endif

if label$ = "fifteen"
        start = Get starting point: 4, j
        end = Get end point: 4, j
        dur = end - start

syllone = Get interval at time: 3, start
syllone$ = Get label of interval: 3, syllone
startsyllone = Get starting point: 3, syllone
endsyllone = Get end point: 3, syllone
syllonedur = endsyllone - startsyllone

sylltwo = Get interval at time: 3, end - 0.001
sylltwo$ = Get label of interval: 3, sylltwo
startsylltwo = Get starting point: 3, sylltwo
endsylltwo = Get end point: 3, sylltwo
sylltwodur = endsylltwo - startsylltwo

durratio = syllonedur/sylltwodur
```

```
# Save to a spreadsheet

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...durratio

endif

if label$ = "sixteen"
        start = Get starting point: 4, j
        end = Get end point: 4, j
        dur = end - start

syllone = Get interval at time: 3, start
syllone$ = Get label of interval: 3, syllone
startsyllone = Get starting point: 3, syllone
endsyllone = Get end point: 3, syllone
syllonedur = endsyllone - startsyllone

sylltwo = Get interval at time: 3, end - 0.001
sylltwo$ = Get label of interval: 3, sylltwo
startsylltwo = Get starting point: 3, sylltwo
endsylltwo = Get end point: 3, sylltwo
sylltwodur = endsylltwo - startsylltwo
```

```
durratio = syllonedur/sylltwodur

# Save to a spreadsheet

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...durratio

endif

if label$ = "eighteen"
    start = Get starting point: 4, j
    end = Get end point: 4, j
    dur = end - start

syllone = Get interval at time: 3, start
syllone$ = Get label of interval: 3, syllone
startsyllone = Get starting point: 3, syllone
endsyllone = Get end point: 3, syllone
syllonedur = endsyllone - startsyllone

sylltwo = Get interval at time: 3, end - 0.001
sylltwo$ = Get label of interval: 3, sylltwo
startsylltwo = Get starting point: 3, sylltwo
endsylltwo = Get end point: 3, sylltwo
sylltwodur = endsylltwo - startsylltwo
```

```
durratio = syllonedur/sylltwodur

# Save to a spreadsheet

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...durratio

endif

if label$ = "nineteen"
        start = Get starting point: 4, j
        end = Get end point: 4, j
        dur = end - start

syllone = Get interval at time: 3, start
syllone$ = Get label of interval: 3, syllone
startsyllone = Get starting point: 3, syllone
endsyllone = Get end point: 3, syllone
syllonedur = endsyllone - startsyllone

sylltwo = Get interval at time: 3, end - 0.001
sylltwo$ = Get label of interval: 3, sylltwo
startsylltwo = Get starting point: 3, sylltwo
endsylltwo = Get end point: 3, sylltwo
```

```
sylltwodur = endsylltwo - startsylltwo

durratio = syllonedur/sylltwodur

# Save to a spreadsheet

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...durratio

endif

endfor

select all
minus Strings list
Remove
endfor

select all
Remove
clearinfo

    appendInfoLine: "Whoo-hoo! It didn't crash!"
```

Appendix G Praat Script to Extract Syllable Intensity (Mean) Ratios

```
writeInfoLine: "Getting intensity ratios..."

# Set directory

directory$ = "/Users/dominiquebranson/Desktop/Test/"

# Create spreadsheet

outputPath$ = "/Users/dominiquebranson/Desktop/Test/Syllable_Int.csv"
writeFileLine: "'outputPath$'", "file, word, int_syllone, int_sylltwo, int_ratio"

# Read all files in the folder

Create Strings as file list... wavlist 'directory$'*.wav
Create Strings as file list... gridlist 'directory$'*.TextGrid
number_files = Get number of strings

for i to number_files
clearinfo

# Create and extract intensity tiers for all sound files

    select Strings wavlist
    current_token$ = Get string... 'i'
```

```
        Read from file... 'directory$''current_token$'
        object_name$ = selected$ ("Sound")
        To Intensity... 50 0.005 1

# Read Text Grid files and extract all intervals on tier 4

        select Strings gridlist
        current_token$ = Get string... 'i'
        Read from file... 'directory$''current_token$'
        nInt = Get number of intervals: 4

# Calculate the intensity for the interval on tier 3

for k from 1 to 'nInt'
        select TextGrid 'object_name$'
        label$ = Get label of interval... 4 'k'
        if label$ = "thirteen"
                start = Get starting point: 4, k
                end = Get end point: 4, k

                        syllone = Get interval at time: 3, start
                        syllone$ = Get label of interval: 3, syllone
                        startsyllone = Get starting point: 3, syllone
                        endsyllone = Get end point: 3, syllone

                        sylltwo = Get interval at time: 3, end - 0.001
                        sylltwo$ = Get label of interval: 3, sylltwo
```

```
                startsylltwo = Get starting point: 3, sylltwo
                endsylltwo = Get end point: 3, sylltwo

        select Intensity 'object_name$'
        meanintsyllone = Get mean... startsyllone endsyllone dB
        meanintsylltwo = Get mean... startsylltwo endsylltwo dB
        intratio = meanintsyllone/meanintsylltwo

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...meanintsyllone, ",",
                        ...meanintsylltwo, ",",
                        ...intratio
        endif

        if label$ = "fourteen"
                start = Get starting point: 4, k
                end = Get end point: 4, k

                        syllone = Get interval at time: 3, start
                        syllone$ = Get label of interval: 3, syllone
                        startsyllone = Get starting point: 3, syllone
                        endsyllone = Get end point: 3, syllone

                        sylltwo = Get interval at time: 3, end - 0.001
                        sylltwo$ = Get label of interval: 3, sylltwo
```

```
                startsylltwo = Get starting point: 3, sylltwo
                endsylltwo = Get end point: 3, sylltwo

        select Intensity 'object_name$'
        meanintsyllone = Get mean... startsyllone endsyllone dB
        meanintsylltwo = Get mean... startsylltwo endsylltwo dB
        intratio = meanintsyllone/meanintsylltwo

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...meanintsyllone, ",",
                        ...meanintsylltwo, ",",
                        ...intratio
        endif

        if label$ = "fifteen"
                start = Get starting point: 4, k
                end = Get end point: 4, k

                        syllone = Get interval at time: 3, start
                        syllone$ = Get label of interval: 3, syllone
                        startsyllone = Get starting point: 3, syllone
                        endsyllone = Get end point: 3, syllone

                        sylltwo = Get interval at time: 3, end - 0.001
                        sylltwo$ = Get label of interval: 3, sylltwo
```

```
                startsylltwo = Get starting point: 3, sylltwo
                endsylltwo = Get end point: 3, sylltwo

        select Intensity 'object_name$'
        meanintsyllone = Get mean... startsyllone endsyllone dB
        meanintsylltwo = Get mean... startsylltwo endsylltwo dB
        intratio = meanintsyllone/meanintsylltwo

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...meanintsyllone, ",",
                        ...meanintsylltwo, ",",
                        ...intratio
        endif

        if label$ = "sixteen"
                start = Get starting point: 4, k
                end = Get end point: 4, k

                        syllone = Get interval at time: 3, start
                        syllone$ = Get label of interval: 3, syllone
                        startsyllone = Get starting point: 3, syllone
                        endsyllone = Get end point: 3, syllone

                        sylltwo = Get interval at time: 3, end - 0.001
                        sylltwo$ = Get label of interval: 3, sylltwo
```

```
                startsylltwo = Get starting point: 3, sylltwo
                endsylltwo = Get end point: 3, sylltwo

        select Intensity 'object_name$'
        meanintsyllone = Get mean... startsyllone endsyllone dB
        meanintsylltwo = Get mean... startsylltwo endsylltwo dB
        intratio = meanintsyllone/meanintsylltwo

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...meanintsyllone, ",",
                        ...meanintsylltwo, ",",
                        ...intratio
        endif

        if label$ = "eighteen"
                start = Get starting point: 4, k
                end = Get end point: 4, k

                        syllone = Get interval at time: 3, start
                        syllone$ = Get label of interval: 3, syllone
                        startsyllone = Get starting point: 3, syllone
                        endsyllone = Get end point: 3, syllone

                        sylltwo = Get interval at time: 3, end - 0.001
                        sylltwo$ = Get label of interval: 3, sylltwo
```

```
                startsylltwo = Get starting point: 3, sylltwo
                endsylltwo = Get end point: 3, sylltwo

        select Intensity 'object_name$'
        meanintsyllone = Get mean... startsyllone endsyllone dB
        meanintsylltwo = Get mean... startsylltwo endsylltwo dB
        intratio = meanintsyllone/meanintsylltwo

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...meanintsyllone, ",",
                        ...meanintsylltwo, ",",
                        ...intratio
        endif

        if label$ = "nineteen"
                start = Get starting point: 4, k
                end = Get end point: 4, k

                        syllone = Get interval at time: 3, start
                        syllone$ = Get label of interval: 3, syllone
                        startsyllone = Get starting point: 3, syllone
                        endsyllone = Get end point: 3, syllone

                        sylltwo = Get interval at time: 3, end - 0.001
                        sylltwo$ = Get label of interval: 3, sylltwo
```

```
                startsylltwo = Get starting point: 3, sylltwo
                endsylltwo = Get end point: 3, sylltwo

        select Intensity 'object_name$'
        meanintsyllone = Get mean... startsyllone endsyllone dB
        meanintsylltwo = Get mean... startsylltwo endsylltwo dB
        intratio = meanintsyllone/meanintsylltwo

appendFileLine: "'outputPath$'",
                        ...object_name$, ",",
                        ...label$, ",",
                        ...meanintsyllone, ",",
                        ...meanintsylltwo, ",",
                        ...intratio
        endif

endfor

endfor

select all
Remove

        appendInfoLine: "Whoo-hoo! It didn't crash!"
```

Appendix H Praat Script to Extracting Formant Values at Five Percent Intervals

See supplemental file "Praat Script for Extracting Formant Values at Five Percent Intervals."

Appendix I Regression Coefficients for Front-Stressing of Initial Syllables Duration/Intensity

See supplemental file "Regression Test Coefficients for Front-Stressing of Initial Syllables Duration_Intensity."

Appendix J Regression Coefficients for Vowel Formant Analyses

See supplemental file "Regression Coefficients for Vowel Formant Analyses."

Bibliography

18 U.S.C. § 249. (2022). https://www.law.cornell.edu/uscode/text/18/249.

Alexander, M. (2020). *The new Jim Crow: Mass incarceration in the age of colorblindness*. The New Press.

Alim, H. S., Rickford, J. R., & Ball, A. F. (2016). Introducing raciolinguistics. *Raciolinguistics: How language shapes our ideas about race*, 1-30.

Athreya, K., & Romero, J. (2015). Land of opportunity: Economic mobility in the United States. *Economic Quarterly*, (2Q), 169-191.

Austen, M. (2020). Production and perception of the PIN-PEN merger. *Journal of Linguistic Geography*, *8*(2), 115-126.

Bailey, G. (1997). When did southern American English begin. *Englishes around the world*, *1*, 255-275.

Baker-Bell, A. (2020). Dismantling anti-black linguistic racism in English language arts classrooms: Toward an anti-racist black language pedagogy. *Theory Into Practice*, *59*(1), 8-21.

Baker-Bell, A. (2020). " We Been Knowin": Toward an Antiracist Language & Literacy Education. *Journal of Language and Literacy Education*, *16*(1), n1.

Baugh, J. (1979). *Linguistic Style-Shifting in Black English*. University of Pennsylvania.

Baugh, J. (1983). A survey of Afro-American English. *Annual Review of Anthropology*, *12*(1), 335-354.

Baugh, J. (1996). Perceptions within a variable paradigm: Black and White racial detection and identification based on speech. *Focus on the USA*, 169-182.

Baugh, J. (2000). *Beyond Ebonics: Linguistic pride and racial prejudice*. Oxford University Press.

Baugh, J. (2005). Linguistic profiling. In *Black linguistics* (pp. 167-180). Routledge.

Baugh, J. (2018). *Linguistics in pursuit of justice*. Cambridge University Press.

Bermeitinger, C. (2016). Priming. In *Psychology and mental Health: Concepts, methodologies, tools, and applications* (pp. 42-88). IGI Global.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, *14*(4), 359-375.

Blackmon, D. A. (2009). *Slavery by another name: The re-enslavement of Black Americans from the Civil War to World War II*. Anchor.

Blumenbach, J. F., & Bendyshe, T. T. (1865). On the natural variety of mankind, 1795.

Bobo, L. D., & Fox, C. (2003). Race, racism, and discrimination: Bridging problems, methods, and theory in social psychological research. *Social psychology quarterly*, *66*(4), 319-332.

Bolnick, D. A. (2008). Individual ancestry inference and the reification of race as a biological phenomenon. *Revisiting race in a genomic age*, 70-85.

Bonilla-Silva, E. (2015). The structure of racism in color-blind,"post-racial" America. *American Behavioral Scientist*, *59*(11), 1358-1376.

Bonilla-Silva, E. (2021). *Racism Without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*. Rowman & Littlefield.

Boyd, Z., Elliott, Z., Fruehwald, J., Hall-Lew, L., & Lawrence, D. (2015, April). An evaluation of sociolinguistic elicitation methods. In *ICPhS*.

Boyles, A. S. (2015). *Race, place, and suburban policing: Too close for comfort*. Univ of California Press.

Braveman, P. A., Arkin, E., Proctor, D., Kauh, T., & Holm, N. (2022). Systemic And Structural Racism: Definitions, Examples, Health Damages, And Approaches To Dismantling: Study examines definitions, examples, health damages, and dismantling systemic and structural racism. *Health Affairs*, *41*(2), 171-178.

Brownstein, M., & Zalta, E. (2019). Implicit bias.

Bucholtz, M. (2003). Sociolinguistic nostalgia and the authentication of identity. *Journal of sociolinguistics*, *7*(3).

Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language and linguistics compass*, *4*(6), 377-389.

Carson, C., & Shepard, K. (2001). *A Call to Conscience: The Landmark Speeches of Dr. Martin Luther King, Jr*. Hachette UK.

Charity, A. H. (2008). African American English: An overview. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse (CLD) Populations*, *15*(2), 33-42.

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *American Economic Review*, *104*(5), 141-147.

Chiang, E. (2017). Post-Racial Ideology and Implicit Racial Bias. *Perspectives*, *9*(1), 3.

Clifford v. Commonwealth of Kentucky. (1999). 97-SC-368-MR

Clopper, C. G. (2004). *Linguistic experience and the perceptual classification of dialect variation*. Indiana University.

Clopper, C. G. (2021). Perception of dialect variation. *The handbook of speech perception*, 333-364.

Corley [Branson], D. (2014). *Sounding Guilty: African American English and Racial Inequality in the Courtroom*. Cornell University.

Criminal justice fact sheet. NAACP. (2022, November 4). https://naacp.org/resources/criminal-justice-fact-sheet

Cukor-Avila, P., & Bailey, G. (2015). Rural Texas African American Vernacular English. *The Oxford handbook of African American language*, 181-200.

Davis, A. Y. (2003). Race and criminalization: Black Americans and the punishment industry. *Criminological perspectives: Essential readings*, *284*.

Dayton, E. (1996). *Grammatical categories of the verb in African-American Vernacular English*. University of Pennsylvania.

Derenoncourt, E. (2022). Can you move to opportunity? Evidence from the Great Migration. *American Economic Review*, *112*(2), 369-408.

Douglass, P., Hesse, B., Nopper, T. K., Thomas, G., & Wun, C. (2016). *Conceptual aphasia in black: Displacing racial formation*. Lexington Books.

Dovidio, J. F., & Gaertner, S. L. (1991). Changes in the expression and assessment of racial prejudice.

Dray, P. (2003). *At the hands of persons unknown: The lynching of Black America*. Modern Library.

Du Bois, W. B. (1910). Reconstruction and its benefits. *The American historical review*, 781-799.

Eades, D. (2010). Sociolinguistics and the legal process. In *Sociolinguistics and the Legal Process*. Multilingual Matters.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: race, crime, and visual processing. *Journal of personality and social psychology, 87*(6), 876.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological science, 17*(5), 383-386.

Farrington, C., King, S., & Kohn, M. (2021). Sources of variation in the speech of African Americans: Perspectives from sociophonetics. *Wiley Interdisciplinary Reviews: Cognitive Science, 12*(3), e1550.

Fasold, R. W. (1972). Tense Marking in Black English. A Linguistic and Social Analysis. Urban Language Series, No. 8.

Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard educational review, 85*(2), 149-171.

Flores-González, N., & Salgado, C. D. (2022). Shifting Racial Schemas: From Post-racial to New "Old-fashioned" Racism. *Sociological Inquiry*, *92*(2), 341-363.

Forrest, J., & Wolfram, W. (2019). The Status Of (ING) in African American LanguageA Quantitative Analysis of Social Factors and Internal Constraints. *American Speech, 94*(1), 72-90.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of phonetics, 34*(4), 409-438.

Foulkes, P., Scobbie, J. M., & Watt, D. (2010). Sociophonetics. *The handbook of phonetic sciences*, 703-754.

Fox, W. S., & Gay, G. (1994). Functions and effects of international teaching assistants. *The Review of Higher Education, 18*(1), 1-24.

Freeman, V., & De Decker, P. (2021). Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *The Journal of the Acoustical Society of America, 149*(2), 1211-1223.

Gaither, S. E., Cohen-Goldberg, A. M., Gidney, C. L., & Maddox, K. B. (2015). Sounding Black or White: Priming identity and biracial speech. *Frontiers in Psychology*, 6, 457.

Garcia, R., Roeser, J., & Kidd, E. (2023). Finding your voice: Voice-specific effects in Tagalog reveal the limits of word order priming. *Cognition, 236*, 105424.

Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American speech*, *69*(1), 30-57.

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of voice, 19*(4), 544-554.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992, March). SWITCHBOARD: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on* (Vol. 1, pp. 517-520). IEEE Computer Society.

Gonzales Rose, J. B. (2019). Color-blind but not color-deaf: Accent discrimination in jury selection. *NYU Rev. L. & Soc. Change, 44*, 309.

Guo, G., Fu, Y., Lee, H., Cai, T., Mullan Harris, K., & Li, Y. (2014). Genetic bio-ancestry and social construction of racial classification in social surveys in the contemporary United States. *Demography*, *51*(1), 141-172.

Green, L. J. (2002). *African American English: a linguistic introduction*. Cambridge University Press.

Gross, K. N. (2006). *Colored Amazons: Crime, Violence, and Black Women in the City of Brotherly Love, 1880–1910*. Duke University Press.

Gruenstein, A., McGraw, I., & Sutherland, A. (2009). A self-transcribing speech corpus: collecting continuous speech with an online educational game. In *International Workshop on Speech and Language Technology in Education*.

Haley, K. (1990). Some complexities in speech identification. *The SECOL Review*, *14*, 101-113.

Harcourt, B. E. (2005). *Illusion of order: The false promise of broken windows policing*. Harvard University Press.

Harpster, T. K., Adams, S. H., & Jarvis, J. P. (2009). Analyzing 911 homicide calls for indicators of guilt or innocence: An exploratory analysis. Homicide Studies, 13(1), 69-93. https://doi.org/10.1177/1088767908328073.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception.

Herman, G. N. (2012). *Plea bargaining*. Juris Publishing, Inc.

Hernández-Campoy, J. M. (2014). Research methods in Sociolinguistics. *Aila Review*, *27*(1), 5-29.

Henderson, T. N. Y. (2016). Property, penality, and (racial) profiling. *Stan. JCR & CL*, *12*, 177.

Hertel, F. R., & Pfeffer, F. T. (2016). The land of opportunity?: trends in social mobility and education in the United States.

Hicks, C. D. (2010). *Talk with you like a woman: African American women, justice, and reform in New York, 1890-1935*. Univ of North Carolina Press.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, *109*(2), 748-763.

Hinton, E., & Cook, D. (2021). The mass criminalization of Black Americans: A historical overview. *Annual Review of Criminology*, *4*, 261-286.

Hinton, L. N., & Pollock, K. E. (2000). Regional variations in the phonological characteristics of African American Vernacular English. *World Englishes*, *19*(1), 59-71.

Holliday, N., Burdin, R., & Tyler, J. (2015). Sandra Bland: talking while black. *Language Log*.

Howell, J., & Korver-Glenn, E. (2018). Neighborhoods, race, and the twenty-first-century housing appraisal industry. *Sociology of Race and Ethnicity*, *4*(4), 473-490.

Hudley, A. H. C., Mallinson, C., & Bucholtz, M. (2020). Toward racial justice in linguistics: Interdisciplinary insights into theorizing race in the discipline and diversifying the profession. *Language*, *96*(4), e200-e235.

Inoue, M. (2006). *Vicarious language: Gender and linguistic modernity in Japan* (Vol. 11). Univ of California Press.

Jefferson-Jones, J., & Henderson, T. N. Y. (2020). # LivingWhileBlack: Blackness as Nuisance.

Johnson, B. D., & Lee, J. G. (2013). Racial disparity under sentencing guidelines: A survey of recent research and emerging perspectives. *Sociology Compass*, *7*(7), 503-514.

Jones, C. P. (2000). Levels of racism: a theoretic framework and a gardener's tale. *American journal of public health*, *90*(8), 1212.

Jones, T., Kalbfeld, J. R., Hancock, R., & Clark, R. (2019). Testifying while black: An experimental study of court reporter accuracy in transcription of African American English. *Language*, *95*(2), e216-e252.

Jost, J., & Ogletree, C. (2011). *The Obamas and a (post) racial America?*. Oxford University Press.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, *28*(4), 441-456.

Kelling, G. L., & Wilson, J. Q. (1982). Broken windows. *Atlantic monthly*, *249*(3), 29-38.

Kelso, J. A. S., Bateson, E.-V., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics and dynamic modeling. Journal of the Acoustical Society of America, 77, 266–280.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation: studies in corpus sociophonetics*. Springer.

Kendall, T., & Farrington, C. (2021). The Corpus of Regional African American Language. Version 2021.07. Eugene, OR: The Online Resources for African American Language Project.

Kendall, T., & Farrington, C. (2022). 14 Managing Sociolinguistic Data with the Corpus of Regional African American Language (CORAAL). *The Open Handbook of Linguistic Data Management*, 185.

Kendall, T., & Thomas, E. R. (2010). Vowels: Vowel manipulation, normalization, and plotting in R. *R package, version*, *1*.

Kendi, I. X. (2023). *How to be an antiracist*. One world.

Kinder, D. R., & Sanders, L. M. (1996). *Divided by color: Racial politics and democratic ideals*. University of Chicago Press.

King, S. (2020). From African American vernacular English to African American language: Rethinking the study of race and language in African Americans' Speech. *Annual Review of Linguistics*, *6*, 285-300.

King, S. (2021). Rethinking race and place: The role of persona in sound change reversal. *Journal of Sociolinguistics*, *25*(2), 159-178.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., ... & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, *117*(14), 7684-7689.

Kovera, M. B. (2019). Racial disparities in the criminal justice system: Prevalence, causes, and a search for solutions. *Journal of Social Issues*, *75*(4), 1139-1164.

Kozol, J., & Perluss, D. (1992). Savage inequalities: Children in America's schools. *Clearinghouse Rev.*, *26*, 398.

Kurinec, C. A., & Weaver III, C. A. (2019). Dialect on trial: use of African American Vernacular English influences juror appraisals. *Psychology, Crime & Law*, *25*(8), 803-828.

Laan, G. P. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*(1), 43-65.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1968). A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City. Volume I: Phonological and Grammatical Analysis.

Labov, W. (1969). A Study of Non-Standard English.

Ladefoged, P., & Johnson, K. (2006). *A course in phonetics*. Cengage learning.

Lane, I., Eck, M., Rottmann, K., & Waibel, A. (2010, June). Tools for collecting speech corpora via Mechanical-Turk. In *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with amazon's mechanical turk* (pp. 184-187).

Lanehart, S. (Ed.). (2015). *The Oxford Handbook of African American Language*. Oxford University Press.

Lanehart, S. (2022). *Language in African American Communities*. Taylor & Francis.

Leemann, A., Jeszenszky, P., Steiner, C., Studerus, M., & Messerli, J. (2020). Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing. *Linguistics Vanguard*, *6*(s3).

Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, *36*(4), 533-554.

Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society*, *43*(5), 539-566.

Lombroso, C. (2006). *Criminal man*. Duke University Press.

Lopez, I. F. H. (1994). The social construction of race: Some observations on illusion, fabrication, and choice. *Harv CR-CLL Rev.*, *29*, 1.

Lum, L. (2009). The Obama era: A post-racial society?. *Diverse Issues in Higher Education*, *25*(26), 14.

MacFarlane, A. E., & Stuart-Smith, J. (2012). 'One of them sounds sort of Glasgow Uni-ish'. Social judgements and fine phonetic variation in Glasgow. *Lingua*, *122*(7), 764-778.

Martin, J. L., & Tang, K. (2020, October). Understanding Racial Disparities in Automatic Speech Recognition: The Case of Habitual" be". In *Interspeech* (pp. 626-630).

Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard university press.

McLarty, J. A. (2019). *Prosodic prominence perception, regional background, ethnicity and experience: Naive perception of African American English and European American English* (Doctoral dissertation, University of Oregon).

McLennan, R. M. (2008). *The crisis of imprisonment: Protest, politics, and the making of the American penal state, 1776-1941*. Cambridge University Press.

Meek, Barbra (2006). And the Injun goes 'How!': Representations of American Indian English in white public space. Language in Society 35(1):93–128.

Mengesha, Z., Heldreth, C., Lahav, M., Sublewski, J., & Tuennerman, E. (2021). "I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans. *Frontiers in Artificial Intelligence*, *4*, 169.

Mishler, M. (2016). *The Atlantic Origins of Mass Incarceration: Punishment, Abolition, and Racial Inequality* (Doctoral dissertation, Ph. D. diss., New York University).

Moberg, S. P., Krysan, M., & Christianson, D. (2019). Racial attitudes in America. *Public Opinion Quarterly*, *83*(2), 450-471.

Molden, D. C. (2014). Understanding priming effects in social psychology: An overview and integration. *Social Cognition*, *32*(Supplement), 243-249.

Montgomery, M., & Eble, C. (2004). Historical perspectives on the pen/pin merger in Southern American English. *Studies in the History of the English Language II*, 415.

Moran, M. J. (1993). Final consonant deletion in African American children speaking Black English: A closer look. *Language, Speech, and Hearing Services in Schools*, *24*(3), 161-166.

Morning, A. (2011). *The nature of race: How scientists think and teach about human difference*. Univ of California Press.

Morning, A. (2014). And you thought we had moved beyond all that: Biological race returns to the social sciences. *Ethnic and Racial Studies*, *37*(10), 1676-1685.

Morrison, T., St John, P., Callahan, J., Callahan, S., & Baker, L. (1975). Black Studies Center Public Dialogue, Part 2.

Nagrani, A., Chung, J. S., & Zisserman, A. (2017). Voxceleb: a large-scale speaker identification dataset. *arXiv preprint arXiv:1706.08612*.

Nash, G. B. (1988). *Forging freedom: The formation of Philadelphia's black community, 1720-1840*. Harvard University Press.

New International Version Bible. (2023). Bible Gateway. https://www.biblegateway.com/.

Neubauer, D. W., & Fradella, H. F. (2018). *America's courts and the criminal justice system*. Cengage Learning.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of language and social psychology*, *18*(1), 62-85.

Oliver, M., & Shapiro, T. (2013). *Black wealth/white wealth: A new perspective on racial inequality*. Routledge.

Omi, M., & Winant, H. (2014). *Racial formation in the United States*. Routledge.

Oshinsky, D. M. (1997). *Worse than slavery*. Simon and Schuster.

Oxford English Dictionary. (2023). https://www.oed.com/

Packard, J. M. (2003). *American nightmare: the history of Jim Crow*. Macmillan.

Pager, D. (2003). The mark of a criminal record. *American journal of sociology*, *108*(5), 937-975.

Paris, D., & Alim, H. S. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard educational review*, *84*(1), 85-100.

Payne, B. K., Vuletich, H. A., & Brown-Iannuzzi, J. L. (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences*, *116*(24), 11693-11698.

Peffley, M., & Hurwitz, J. (1998). Whites' stereotypes of Blacks: Sources and political consequences. *Perception and prejudice: Race and politics in the United States*, 58-99.

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition*, *114*(1), 42-55.

Pew Research Center. (2021, April 7). *Internet/broadband fact sheet*. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Poorjam, A. H., Little, M. A., Jensen, J. R., & Christensen, M. G. (2019). Quality control in remote speech data collection. *IEEE Journal of Selected Topics in Signal Processing*, *13*(2), 236-243.

Prolific. Prolific · Quickly find research participants you can trust. (n.d.). http://www.prolific.co/

Provine, D. M. (2011). Race and inequality in the war on drugs. *Annual Review of Law and Social Science*, *7*, 41-60.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of language and social psychology*, *18*(1), 10-30.

Pyke, K. D. (2010). What is internalized racial oppression and why don't we study it? Acknowledging racism's hidden injuries. *Sociological perspectives*, *53*(4), 551-572.

Qualtrics XM. Qualtrics. (2023, June 14). https://www.qualtrics.com/

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rawat, S., Gupta, P., & Kumar, P. (2014, November). Digital life assistant using automated speech recognition. In *2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH)* (pp. 43-47). IEEE.

Rehavi, M. M., & Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy*, *122*(6), 1320-1354.

Rickford, J. R. (1999). The Ebonics controversy in my backyard: A sociolinguist's experiences and reflections. *Journal of Sociolinguistics*, *3*(2), 267-275.

Rickford, J. R., Sweetland, J., & Rickford, A. E. (2004). African American English and other vernaculars in education: A topic-coded bibliography. *Journal of English Linguistics*, *32*(3), 230-320.

Rickford, J. R. (2005). Using the vernacular to teach the standard. *Ebonics: The urban education debate*, *2*, 18-40.

Rickford, J. R., & King, S. (2016). Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language*, 948-988.

Robinson, K. L. (2022). Saving the Black Child: Dozier School for Boys and the Racialization of Juvenile Justice. *Spectrum: A Journal on Black Men*, *10*(1), 37-56.

Rosa, J., & Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in society*, *46*(5), 621-647.

Rothman, D. J. (2017). *The discovery of the asylum: Social order and disorder in the new republic*. Routledge.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher education*, *33*, 511-531.

Schleiden, C., Soloski, K. L., Milstead, K., & Rhynehart, A. (2020). Racial disparities in arrests: a race specific model explaining arrest rates across black and white young adults. *Child and adolescent social work journal*, *37*, 1-14.

Schrader, S. (2019). *Badges without borders: How global counterinsurgency transformed American policing* (Vol. 56). University of California Press.

Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations*. Harvard University Press.

Schwarz, P. J. (1998). *Twice condemned: Slaves and the criminal laws of Virginia, 1705-1865*. The Lawbook Exchange, Ltd.

Serre, D., & Pääbo, S. (2004). Evidence for gradients of human genetic diversity within and among continents. *Genome research*, *14*(9), 1679-1685.

Shapiro, T. M. (2004). *The hidden cost of being African American: How wealth perpetuates inequality*. Oxford University Press, USA.

Sheatsley, P. B. (1966). White attitudes toward the Negro. *Daedalus*, 217-238.

Shiao, J. L., Bode, T., Beyer, A., & Selvig, D. (2012). The genomic challenge to the social construction of race. *Sociological Theory*, *30*(2), 67-88.

Simpson, A. P. (2013). Spontaneous speech. *The Bloomsbury companion to phonetics*, 155-169.

Sirin, C. V. (2011). From Nixon's War on Drugs to Obama's drug policies today: Presidential progress in addressing racial injustices and disparities. *Race, Gender & Class*, 82-99.

Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *American psychologist*, *60*(1), 16.

Smiley, C., & Fakunle, D. (2016). From "brute" to "thug:" The demonization and criminalization of unarmed Black male victims in America. *Journal of human behavior in the social environment*, *26*(3-4), 350-366.

Smitherman, G. (1977). *Talkin' and Testifyin': The Language of Black America*. Boston: Houghton Mifflin.

Smitherman, G. (1988). Discriminatory discourse on Afro-American speech. *Discourse and discrimination*, 144-175.

Sommers, S. R., & Marotta, S. A. (2014). Racial disparities in legal outcomes: On policing, charging decisions, and criminal trial proceedings. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 103-111.

Spears, A. K. (1982). The Black English semi-auxiliary come. *Language*, 850-872.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, *79*(4), 1086-1100.

Taylor, A. A. (1938). Historians of the Reconstruction. *The Journal of Negro History*, *23*(1), 16-34.

Tesler, M. (2020). *Post-racial or most-racial? Race and politics in the Obama era*. University of Chicago Press.

Tesler, M., & Sears, D. O. (2010). Obama's Race: The 2008 Election and the Dream of a Post Racial America.

Thomas, E. R. (2007). Phonological and phonetic characteristics of African American vernacular English. *Language and Linguistics Compass*, *1*(5), 450-475.

Thomas, E. R. (2015). Prosodic features of African American English. *The Oxford handbook of African American Language*, 420-438.

Thomas, E. R. (2020). Sociophonetic trends in studies of Southern US English. *The Journal of the Acoustical Society of America*, *147*(1), 529-540.

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite.

Thomas, E. R., & Reaser, J. (2004). Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. *Journal of sociolinguistics*, *8*(1), 54-87.

Thurston, B. (n.d.). *Living While Black Headlines*. Airtable. https://airtable.com/shrq7t6zz5zwImq54/tblftqCxO7gcK1Ol3/viwAMSwE7vNXtV1nR?blocks=hide

Traunmüller, H. (1997). Auditory scales of frequency representation.

U.S. Census Bureau. (2021). *2017 – 2021 American Community Survey 5-year Selected Population Tables.* Retrieved from https://www.census.gov/programs-surveys/acs/data.html

Virtanen, S. V., & Huddy, L. (1998). Old-fashioned racism and new forms of racial prejudice. *The journal of politics*, *60*(2), 311-332.

Voice and video surveys. Phonic. (n.d.). http://www.phonic.ai/

Wagner, B. (2009). *Disturbing the peace: Black culture and the police power after slavery*. Harvard University Press.

Walker, S. (1998). *Popular justice: A history of American criminal justice*. Oxford University Press, USA.

Weissler, R. E. (2021). *Leveraging African American English Knowledge: Cognition and Multidialectal Processing* (Doctoral dissertation).

Weissler, R. E. (2022). A meeting of the minds: Broadening horizons in the study of linguistic discrimination and social justice through sociolinguistic and psycholinguistic approaches. *Annual Review of Applied Linguistics*, 1-7.

Weldon, T. L. (2021). *Middle-class African American English*. Cambridge University Press.

Weldon, T. L., & Moody, S. (2015). The place of Gullah in the African American linguistic continuum. *The Oxford handbook of African American language*, 163-180.

Wells, I. B. (1991). A Red Record. 1895. *Selected Works of Ida B. Wells-Barnett. Intro. Trudier Harris. New York: Oxford UP*, 138-252.

Wolfram, W. A. (1969). A Sociolinguistic Description of Detroit Negro Speech. Urban Language Series, No. 5.

Wolfram, W. (2004). The grammar of urban African American vernacular English. *Handbook of varieties of English*, *2*, 111-32.

Wolfram, W. (2007). Sociolinguistic folklore in the study of African American English. *Language and Linguistics Compass*, *1*(4), 292-313.

Wolfram, W., & Thomas, E.R. (2002). *The development of African American English*. Oxford: Blackwell.

Wright, K. E. (2017). The Reflection and Reification of Racialized Language in Popular Media.

Wright, K. E. (2019). *Experiments on Linguistic Profiling of Three American Dialects* (Doctoral dissertation).

Yang, C. S. (2015). Free at last? Judicial discretion and racial disparities in federal sentencing. *The Journal of Legal Studies*, *44*(1), 75-111.