Ideology, Awareness, and Sociophonetic Perception in Asian American and Canadian Speech

by

Lauretta S. P. Cheng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2024

Doctoral Committee:

Professor Andries W. Coetzee, Co-chair
Associate Professor Jonathan R. Brennan, Co-chair
Associate Professor Annette D'Onofrio, Northwestern University
Professor Susan Gelman
Assistant Professor Savithry Namboodiripad

Lauretta S. P. Cheng

lspcheng@umich.edu

ORCID iD: 0000-0003-1992-6594

ACKNOWLEDGEMENTS

I would first like to thank my dissertation committee: Jonathan Brennan, Andries Coetzee, Annette D'Onofrio, Susan Gelman, and Savithry Namboodiripad. All the time, energy, feedback, and encouragement they gave to me throughout this process has been invaluable, to both the completion of this dissertation and to my growth as a researcher. I am immensely grateful to my co-advisors, Andries and Jon, who spent countless hours with me in joint advising meetings and who were always supportive of whatever I chose to pursue and whatever decisions I made. I was able to make it to the end of this journey because of their consistent guidance and insight—thank you, both. Annette's work inspired much of this dissertation, and I benefited enormously from the perspective, advice, and excitement she provided at every meeting we had. I am thankful to Susan who, despite her different area of expertise, always expressed interest in my work and brought such positive energy to our interactions. I am particularly indebted to Savi, whose way of thinking, doing research, and mentoring has greatly shaped my own approaches. Her interdisciplinary knowledge, enthusiasm, and generosity are unmatched, and I feel lucky to have worked closely with her as a collaborator. Thank you for being such a relatable and supportive teacher and mentor.

In addition to my committee members, this dissertation was made possible by the contributions of too many people to list. I am grateful for every person who took even a moment to help me forward a recruitment email. I'd especially like to thank those who provided inspiration for the earliest ideas; offered suggestions at various stages of the project; as well as assisted with creating materials, recruiting participants, and processing data: Mathew Jacqmin-Kramer, Andrew Cheng, Wilkinson Gonzales, Pocholo Umbal, Mecca Petross, Aarya Menon, Diane Yu, Rachel Paroff, Noah Luntzlara, Rachel Soo, Susan Lin, Chantal De Leon, Sylvia Cho, Daisy Choi, Karen Tsai, Anna Alsop, Naomi Cui, Irene Yi, Cynthia Zhong, Connie Ting, Katherine Sung, Michelle Kamigaki-Baron, Stephanie Chung, Yu Cen Gao, Kaining Xu, Stacy Liu, Hayley Heaton, Jian Zhu, Dominque Canning, Danielle Burgess, Felicia Bisnath, Emma Portugal, Jeonghwa Cho, Tzu-Yun Tung, Yourdanis Sedarous, Alicia Stevers, Moira Saltzman, and Molly Babel. Thank you to all of the people who gave their time to participate in my studies, and to all of those who were excited about this research and saw value in it—those conversations energized me and kept me going.

I have been impacted by so many others in the U-M Linguistics community in my time here. I want to express my gratitude to the many faculty that I've learned from in a variety of contexts: Pam Beddor, Robin Queen, Steve Abney, Marlyse Baptista, Lisa Levinson, Acrisio Pires, Jelena Krivokapic, Natasha Abner, Ezra Keshet, San Duanmu, Sally Thomason. Thank you to everyone who provided me with community, support, and feedback, including in the Contact, Cognition, and Change Lab; the Computational Neurolinguistics Lab; and the SoConDi and PhonDi discussion groups. I'd also like to thank Jennifer Nguyen, Carl Abrego, and Elexis Nelson for smoothly handling my many administrative questions and requests over the years. I am grateful to have been funded in part by the Social Sciences and Humanities Research Council of Canada through the SSHRC Doctoral Fellowship, along with generous funding from the Rackham Graduate School and Linguistics Department at U-M to conduct and disseminate my research.

Many thanks to all of the Linguistics graduate students I learned with, taught with, worked with, chatted with, and laughed with in the grad lounge, offices,

and hallways of Lorch Hall (and then over Zoom), including: Aliaksei Akimenka, Danuta Allen, Wyatt Barnes, Felicia Bisnath, Rawan Bonais, Dominique Bouavichith, Danielle Burgess, Ian Calloway, Dominique Canning, Lucy Yu-Chuan Chiang, Jeonghwa Cho, Justin Craft, Sophia Eakins, Wilkinson Daniel Wong Gonzales, Tamarae Hildebrandt, Marjorie Herbert, Hayley Heaton, Mathew Jacqmin-Kramer, Demet Kayabaşi, Jiseung Kim, Chia-Wen Lo, Andrew McInnerney, Carrie Ann Morgan, L.R. "Nik" Nikolai, Joy Peltier, Emma Portugal, Emily Sabo, Jungyun Seo, Moira Saltzman, Yourdanis Sedarous, Kate Sherwood, Cecilia Solís-Barroso, Yushi Sugimoto, Csilla Tatár, Tzu-Yun Tung, Rachel Weissler, Anna Whitney, Kelly Wright, Junyuan Zhao, and Jian Zhu. I especially want to thank Jiseung, Tzu-Yun, Jeonghwa, Lucy, Dani, Dominique, Yourdanis, and Felicia for your company, support, and advice over the years.

To my cohort, Mathew, Wil, and Yushi: My time at U-M would not have been half as enjoyable or memorable without you three. I learned and gained so much from each of you. Thank you for all the time we spent together, inside and outside of Lorch Hall, from cleaning the grad lounge on the first day to attending each others' defenses. I'm glad you were the ones I went through this program with.

Mathew, it would be impossible to enumerate all the ways you supported and continue to support me, in both practical matters and emotional matters. Thank you for being there for me, for listening, for letting me be myself. Thank you for sharing, learning, and growing with me. You've helped to broaden my perspective in so many ways. This work definitely would not be what it is today without you. I also have to thank our cats, Miso and Tofu, who provided me with comfort, company, joy, and laughter to get me through each day.

To my family and friends back home, thank you for welcoming me back each time as if I had never left. Especially, thank you to my parents whose support has allowed me to come all this way to the end of this long path.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

Appendix

ABSTRACT

Social information is cognitively linked to linguistic information, evidenced by bidirectional influences on perceptual processing of speech. Models of sociophonetic cognition theorize that the way linguistic experiences are interpreted and stored in memory is mediated by listener attention, which is guided by ideology. This relationship, however, has not yet been tested in depth, especially at the level of the individual. In this dissertation, I focus on two constructs related to linguistic ideologies—*personae* (recognizable social types) and *awareness* (conscious access to sociolinguistic variation)—exploring their role in perceptual cognitive representations of sociophonetic variation.

The sociolinguistic context examined in this dissertation is Asian American and Canadian (AAC) speech, where AAC is approached as a panethnic ideological construct tied to macrosocial Asian-racialized ideologies in the minds of listeners. Representations of AACs may be further mediated by intraethnic categories at a microsocial level. Despite the sizable and growing presence of AACs in North America, linguistic variation in AAC speech is little understood, particularly from the perspective of listeners. This dissertation explores the ideological landscape of AAC speech through the lens of personae and awareness, documenting AAC-associated sociophonetic variation and its connection to cognitive representations and perceptual processing of AAC speech.

To test the influence of ideology on sociophonetic cognition and enrich descriptions of AAC speech, I examine production, perception, and metalinguistic data from Californian Asian Americans across four studies. I begin with more naturalistic and exploratory approaches followed by controlled and targeted methods. First, a corpus analysis of YouTube-sourced speech establishes /oʊ/-backing as a phonetic feature found in many AACs' production. Second, a perceptual task based on the YouTube corpus along with a metalinguistic survey indicate that /oʊ/-backing and /ð/-stopping are ideological features of AAC speech that AAC listeners can consciously access and use to make perceptual ethnic identification decisions. Third, open-ended comments and social evaluation surveys demonstrate that AACs can be ideologically subdivided into microsocial personae representing a more culturally Asian (or ethnically-oriented) AAC type compared to a more bicultural (or mainstream-oriented) AAC type, each being tied to a cluster of visual and linguistic features. Finally, a perceptual social categorization task paired with gradient assessments of awareness reveal that hearing /ð/-stopping—a variant with higher awareness—is strongly associated with an ethnically-oriented AAC persona while hearing /oʊ/-backing—a variant with lower awareness—is weakly associated with both ethnically- and mainstream-oriented AAC personae.

The results show that accounting for personae and awareness contributes to a fuller, more realistic characterization of AAC speech and how it is processed. Moreover, they align with an interpretation where personae and awareness play a mediating role in sociophonetic representations and perceptual processing with regards to the nature of social information and the strength of sociophonetic associations. On the whole, this dissertation highlights the relevance of ideology both in the context of AAC speech and in generalized models of sociolinguistic cognition, supporting moves towards theoretical and methodological approaches that foreground ideology to understand how speech is processed in the social world.

CHAPTER I

Introduction

This dissertation broadly examines the interaction between socioindexical and linguistic knowledge through the case of Asian American and Canadian speech. In particular, I seek to explore the role of ideology on the encoding and perceptual processing of sociophonetic variation, speaking to both the nature of sociophonetic representations and how they may be learned.

A plethora of evidence indicates that social and linguistic knowledge are cognitively linked (e.g., Pierrehumbert, 2016; Hay, 2018), affecting not only speech production (e.g., linguistic variants produced by speakers) but also speech perception (e.g., linguistic variants expected and perceived by listeners). Listeners' sociophonetic representations and perceptual processing are shaped by factors beyond simply amount of linguistic input (Sumner et al., 2014); this suggests that linguistic experiences are crucially mediated by attention, influencing how they are perceived and stored in memory (e.g., Pierrehumbert, 2006; G. J. Docherty & Foulkes, 2014). Such an approach highlights the importance of the individual, including the particular linguistic ideologies they hold that guide their attention and determine the social information that is ultimately indexed to phonetic detail. In addition to macrosocial categories, associated social information may come in the form of social personae, ideological constructs representing holistic social types of people (e.g., D'Onofrio, 2020). On the whole, attention-weighting exemplar theoretic approaches argue that ideology and awareness of those ideologies—theorized to relate to attention via salience—should be important factors in this picture of sociophonetic cognition (Drager & Kirtley, 2016).

I examine these issues in the context of Asian American and Canadian (AAC) speech, approaching "Asian American" or "Asian Canadian" as a pan-ethnic ideological construct of individuals of Asian origin in the US and Canada (e.g., Kibria, 1998; J. Lee, 2019). Linguistic ideologies about AAC speech intersect with racialized social stereotypes (e.g., "forever foreigner" and "honorary white", Tuan, 1998) and speech styles (e.g., "Mock Asian", Chun, 2004). Despite implications for discrimination faced by AACs who make up a rapidly growing segment of the North American population, linguistic variation in AAC speech and the related linguistic ideologies remain understudied (Reyes & Lo, 2009). Consequently, current understanding of the sociolinguistic expectations and perceptual processing of AAC speech is also limited (e.g., Newman & Wu, 2011; D'Onofrio, 2019). I take this opportunity to further probe various facets of the AAC sociolinguistic context, including linguistic features indexed to AAC identity, social personae representing types of AACs, and the awareness associated with those ideologies. Then, I build on these findings to ask how listeners' ideologies and awareness influence cognitive representations and perceptual processing of AAC speech.

The goals of this dissertation are two-fold. First, I aim to investigate the influence of ideology on sociophonetic representations, perception, and processing, testing predictions based on attention-weighting exemplar theory. Second, I aim to contribute empirical evidence from production, perception, and metalinguistic commentary to our understanding of AAC speech, with implications for how AACs are socially and linguistically perceived by others. My approach will focus on insights gained from individual-centered analysis of speakers and listeners.

Altogether, this work seeks to foreground the social and linguistic experiences of AACs that are often erased in both academic research and broader society. The three guiding questions of this dissertation are as follows:

1. How is AAC speech ideologically constructed, cognitively represented, and perceptually processed?
2. How do social personae factor into sociophonetic representations and perceptual processing?
3. Does sociolinguistic awareness mediate the strength of sociophonetic representations and perceptual processing?

1.1 Outline of Chapters

In Chapter II, I present my theoretical framework, providing an overview of the background literature with regards to sociophonetic representations and processing, followed by a brief introduction to the AAC context. This includes a review of attention-weighting exemplar theoretic models and synthesis of literature from multiple traditions to conceptualize the key constructs of ideology, awareness, and salience.

The next two chapters (Chapter III and IV) review the literature on AAC speech in more depth and present exploratory studies that aim to further establish specific descriptive expectations of AAC speech with regards to phonetic features and social types. Chapter III begins with a review of what we know thus far about potential phonetic features of AAC speech examined in production. Then, I detail a corpus production study of Asian American-identified speakers based on YouTube videos, examining three potential features of AAC speech: /oʊ/-backing, /oʊ/-monophthongization, and speech rhythm.

In Chapter IV, I delve into literature on potential cues to AAC ethnic identification in perception, as well as overarching linguistic ideologies relating to AAC speech. Using naturalistic speech samples from the same YouTube videos as Chapter III, I follow up on the production findings with a study investigating /oʊ/-backing as a phonetic feature linked to AAC identity via assessments of perceptual behavior and ideological self-report. This study serves three distinct purposes, setting the stage for the remainder of the dissertation: (a) examining Asian ethnic identification and the potential role of /oʊ/-backing; (b) quantifying metalinguistic awareness of various potential features of AAC speech, including /oʊ/-backing; and (c) exploring ideological commentary about potential social types of AACs and their links to speech style.

In addition to further clarifying how AAC speech is represented and processed, the following pair of chapters (Chapter V and VI) seek to contribute to existing literature on sociolinguistic cognition by examining how personae and awareness contribute to sociophonetic representations and modulate perceptual processing. In Chapter V, I report on a series of three social evaluation surveys involving AAC(-perceived) faces and voices that add to our understanding of the relevant ideologies involved in representations of AAC speech, specifically as they relate to AAC personae differing along a continuum of acculturation.

Building on the preceding chapters, Chapter VI presents a social categorization experiment testing the role of personae and awareness while further assessing how the specific linguistic variables of /oʊ/-backing and /ð/-stopping may be associated with variable degrees of awareness and linked to microsocial persona-based categories of AACs. The study seeks to solidify our understanding of /oʊ/-backing and /ð/-stopping as sociophonetic variables indexed to AAC identities and provide evidence for the import of personae and awareness in models of sociophonetic cognition. Finally, in Chapter VII, I synthesize the content across all chapters and detail possible directions for future research.

CHAPTER II

Theoretical Framework

2.1 Sociophonetic Representation and Perceptual Processing

Language is inherently social. Both socioindexical and linguistic knowledge interact with the other, where knowledge refers to any kind of cognitive representation including linguistic and metalinguistic memories, beliefs, and attitudes. Although speech perception research in the past framed phonetic variation as a "problem" for listeners to overcome, more recent work discusses variation— including sociophonetic variation—as a part of facilitating speech processing. Under this view, listeners use their stored knowledge about structured linguistic variation, gained through experience, to inform their perceptual processing. Those memory representations and later perception are influenced by various factors, including exposure frequency but also beliefs and ideologies, including social stereotypes; ultimately, it is subjective interpreted experiences of events, episodes, or instances—rather than objective "actual" experiences—that influence memory and perception (Smith & Zárate, 1992; G. J. Docherty & Foulkes, 2014; Sumner et al., 2014). Sociolinguistic expectations have been demonstrated to modulate linguistic perception many times over (see Campbell-Kibler, 2010b; Drager, 2010). Still, the full extent of how perceptual processing incorporates social information is yet to be understood. For example, open questions remain as to how sociophonetic associations of different types are acquired, represented, and accessed during speech processing (e.g., G. J. Docherty & Foulkes, 2014; Drager & Kirtley, 2016; D'Onofrio, 2018a). Taking this as a starting point, I focus my attention on nuances of how an individual perceiver's experience influences perceptual cognition and behavior.

The current research also broadly engages with two unfolding research trends. Many scholars have been moving towards the integration of social, linguistic, and cognitive fields to better understand these complex questions (Chevrot et al., 2018); this includes the area of *sociolinguistic cognition* (Campbell-Kibler, 2010b, 2016) which involves methods from psycholinguistics and social cognition to "explor[e] the cognitive and cerebral mechanisms underpinning the ability to encode sociolinguistic variation, to implement it during speech production, and to process it during speech perception" (p. 685). Campbell-Kibler (2016) advocates for engagement with neighbouring fields of study, particularly social cognition and cognitive psychology, to approach sociolinguistic behavior (including constructs such as awareness and salience) in cognitively realistic ways. In line with this approach, I draw on various traditions to inform my theoretical framework and more fully ground sociophonetic theorizing in various aspects of cognition.

In addition, there is an increasing focus on analysis at the level of individuals to understand broader, community patterns of linguistic behavior, including in areas of study such as sound change, phonetics, and phonology (Yao & Chang, 2016; Coetzee, 2018; Yu & Zellou, 2019), third wave sociolinguistics (Eckert, 2008a, 2012; D'Onofrio, 2020), and sociophonetics and speech perception (Foulkes, 2010; G. J. Docherty & Foulkes, 2014). These approaches recognize that all community patterns are emergent over individuals, thus highlighting the crucial role of individual cognition to elucidate not only how language processing works, but also how sociolinguistic meanings are acquired and how innovations initiate, and then spread at community levels. Accordingly, I use the individual as a lens to further probe how socioindexical associations form and influence perceptual processing.

In the following, I provide a review of the relevant background literature on sociophonetic perception, theoretical approaches to perceptual processing, and the key concepts to be explored (e.g., awareness, salience), as well as targeted research topics in need of further development.

2.1.1 Sociophonetic Perception

Accumulating evidence in speech perception and sociophonetics suggests a complex, intertwined, and bidirectional relationship of socioindexical and linguistic information (Campbell-Kibler, 2010a; Drager, 2010; G. J. Docherty & Foulkes, 2014; Sumner et al., 2014; Bouavichith et al., 2019). On the one hand, phonetic information can convey hints about speaker identity as well as influencing social impressions or evaluations of the speaker. For example, listeners have perceptual representations of regional varieties and can consistently, though not always accurately, provide judgments on a speaker's regional origin, particularly for local varieties (e.g., Clopper & Pisoni, 2004) and stereotyped varieties (e.g., Clopper & Pisoni, 2006, for a review, see Clopper, 2021). Similarly, listeners can identify a talkers' ethnic group affiliation, such as differentiating African Americans from European Americans, at rates above chance (Thomas & Reaser, 2004), even with very short samples of speech, indicating that this extraction of social information is extremely rapid (Purnell et al., 1999; Scharinger et al., 2011). Listeners also make relatively consistent judgments about talkers' social characteristics, complexly related to particular combinations of phonetic cues, voices, and contexts (e.g., Campbell-Kibler, 2007; Montgomery & Moore, 2018; Drager et al., 2021). Findings such as these are tied to social consequences of linguistic discrimination and injustice, particularly for marginalized speakers (Purnell et al., 1999; Craft et al., 2020).

Besides social interpretations sourced from language cues, there are linguistic consequences of social factors; this is the main focus of this section. Social inferences made via prior speech, or based on external signals such as visual information (e.g., a face) and other contextual information (e.g., how the speaker is described), can influence how later speech is interpreted. Social priming paradigms, whereby such information is provided prior to and/or concurrent with the linguistic stimuli, is commonly used to test the effect of socioindexical information on the processing of linguistic information. Within this strand of research, many studies indicate that social categories such as gender (Strand, 1999; Bouavichith et al., 2019), age (Koops et al., 2008; Drager, 2011), social class (Hay, Warren, et al., 2006), race/ethnicity (Staum Casasanto, 2010; McGowan & Babel, 2019), and regional or national origin (Niedzielski, 1999; Hay, Nolan, et al., 2006) impact speech perception outcomes.

These broadly-characterized effects of social information on linguistic outcomes are found across many different tasks, response types, and timings. For example, some studies find evidence of shifted "low-level" phoneme interpretation, such as perceiving an ambiguous token as [s] or [ʃ] (e.g., Strand, 1999) or showing more or less discrimination between vowels (e.g., McGowan & Babel, 2019), while others report improved or worsened performance in "high-level" sentence intelligibility tasks, being more or less accurate in identifying words embedded in noise (e.g., M. Babel & Russell, 2015; McGowan, 2015). That is, some social priming effects present as altered perception while others appear as so-called processing costs or facilitators. In addition, supporting evidence has been found across both early and late timings, using measures of early eye movements (e.g., Koops et al., 2008; D'Onofrio, 2018a) or neural responses (e.g., Loudermilk, 2013), as well as untimed behavioral responses (e.g., Hay, Warren, et al., 2006) and delayed declarative memory tasks (D'Onofrio, 2021b). On the whole, this sizable body of literature demonstrates a relationship between listeners' expectations of a talker and their perceptual behavior, though the underlying mechanisms or processes are unclear. A closer look at the evidence provides hints as to how these perceptual outcomes result, in terms of both the cognitive representations that contribute to expectations and in-the-moment cognitive processes.

First, findings indicate that ideological categories or stereotypes influence outcomes like perceptual responses, lexical access, and memory recall—that is, beliefs about how certain groups of people speak, whether or not that is truly reflective of those groups, impact behavior (Niedzielski, 1999; Hay, Nolan, et al., 2006; J. Kim, 2016; J. Kim & Drager, 2018; McGowan & Babel, 2019). For example, the overtly stereotyped notion that Canadians but not Michiganders pronounce /aʊ/ with a raised nucleus preceding voiceless obstruents, as in "about", was reported to induce shifted perceptual responses by Michigan listeners (Niedzielski, 1999). Relatedly, despite how it is often discussed, the social information that listeners use to form expectations can be rather nuanced, reflecting detailed and/or intersectional social constructs. For example, *personae*—described as "holistic, ideological social types that are recognizably linked with ways of being and speaking" (D'Onofrio, 2020)—such as "Valley girl" or "business professional" can influence perceptual behavior in different ways (D'Onofrio, 2018b, 2019, 2021b). In other words, sociolinguistic expectations are not based only on direct, interactional speech input (e.g., resulting in effects of frequency or recency) or broad macrosocial categories (e.g., leading to effects of age, gender, or ethnicity). While sociolinguistic associations may or may not be grounded in typically-occurring linguistic input, they nevertheless arise from listeners' experiences with language and language ideology to influence their later perceptual processing.

Further, degree of awareness (for the moment used interchangeably with salience) seems to play a role. While evidence indicates that social priming results certainly can occur, effects are in fact variably found across individuals and groups such that some people show no priming at all (e.g., Drager, 2011; Loudermilk, 2013; Walker et al., 2019; Alderton, 2020), suggested in some cases to be tied to salience or awareness of the relevant ideological associations. The reasoning is that differences in experiences result in varying levels of salience, which then modulate strength of expectations and resulting perception. In this vein, a few studies have examined the role of awareness or salience on perceptual outcomes. Although some studies do not find such an effect, they tended to not directly or gradiently measure degree of sociolinguistic awareness/salience (e.g., Lawrence, 2015; Chang, 2017). Other findings suggest that linguistic variables that are higher in community-level salience (Juskan, 2018) or individuals who report higher awareness of a particular stereotype (L. S. P. Cheng, 2021) may lead to a greater degree of social priming on perception. These results highlight the importance of nuance—as in the case of personae rather than macrosocial categories—in considering not only the existence but also the degree of awareness or salience, and how this affects representation and expectations.

Notably, however, the results of social priming appear not to depend on conscious beliefs or attributions *at the time of perception* (Hay, Nolan, et al., 2006; Hay & Drager, 2010; McGowan & Babel, 2019). In one example, listeners explicitly reported that they did not believe the prime manipulation suggesting that the talker was Australian, but group-level behavioral results nevertheless showed evidence of priming effects (Hay, Nolan, et al., 2006). In another case where the same voice was presented across two blocks accompanied by different social information, we find the reverse situation such that listeners claimed to believe the new prime manipulation but overall did not demonstrate priming effects (McGowan & Babel, 2019). Moreover, some studies indicate that the activation of social category concepts due to environmental cues (e.g., stuffed animals) is enough to trigger social priming effects even in the absence of attribution of social information to speakers

(Hay & Drager, 2010, but see Walker et al., 2019). These findings suggest that, even if listeners are consciously aware of a particular association or stereotype, the in-the-moment influence of social information on linguistic processing can operate at an unconscious and/or automatic level, and may persist across time.

It appears that the kinds of sociolinguistic categories or associations in listeners' representations, their degree of awareness, and aspects of processing such as automaticity or timing may be relevant components of sociophonetic perception that have yet to be fully explored. Probing these factors further could help to clarify inner workings of the process of sociophonetic perception. Along these lines, work by D'Onofrio highlights some outstanding issues at the intersection of these factors. Examining perception of TRAP-backing, a sociolinguistic variable with various documented social associations, D'Onofrio (2018a) finds that macrosocial (California) speaker information influenced early eye movements to TRAP-backing words while persona (Valley girl) speaker information influenced later behavioral responses. These results could have various interpretations. Processing across timing may vary based on the scope of social representations (e.g., broader macrosocial categories are accessed earlier than narrower personae categories). Alternatively, the difference in timing may be conditioned by type of representation—linked to awareness—based on its source or history. For example, are unconscious, directly-acquired, interaction-based associations accessed at early, automatic stages of processing while conscious, indirectly-acquired, ideology-based stereotypes are activated at later, more controlled stages of processing? Given these possibilities, D'Onofrio identifies a need for more research that "explore[s] the many means by which implicit sociolinguistic expectations can be created and how this bears on different stages of sociolinguistic perception" (p. 282).

In all, this literature leaves open questions about types of social categories and how they influence speech perception, as well as the role of awareness and salience. Theorization about this set of results is often discussed and framed within exemplar theory, which I turn to next; using this framework, researchers have proposed how stereotypes, awareness, and salience relate to sociophonetic perception.

2.1.2 Approaches to Representation and Processing

2.1.2.1 Exemplar Theory

Exemplar theory (e.g., Johnson et al., 1997; Pierrehumbert, 2001; Foulkes & Docherty, 2006; Pierrehumbert, 2006, 2016) constitutes the predominant approach in sociophonetic research to account for the numerous results showing that listeners incorporate socioindexical information when both producing and perceiving speech. With its origins as a model of memory in psychology and cognition (e.g., Nosofsky, 1989; see also social cognition, Smith & Zárate, 1992), exemplar-based (and more broadly instance-based) models assume that specific, detailed knowledge (as opposed to abstract or generalized knowledge, as in prototypes) is retained in memory and implicitly affects later behavior. Beyond speech perception, this approach has also gained ground in various other areas of linguistics (for recent treatments, see Ambridge, 2020a, 2020b, on exemplars across domains from an acquisition perspective, Jones, 2019 on perspectives from semantics and natural language processing).

A basic exemplar theory approach to speech perception can be summarized as follows. *Exemplars* are cognitive representations of episodes, instances, or traces of experiences stored in long-term memory (often conceptualized at the level of words; Hay, 2018). They are richly encoded with phonetic and contextual details (linguistic, social, environmental), though this is crucially mediated by attention of the perceiver (e.g., Pierrehumbert, 2006; G. J. Docherty & Foulkes, 2014; Sumner et al., 2014). The accumulation of exemplars over time leads to cognitive structuring of similar exemplars (e.g. clusters and sub-clusters), allowing for emergent higher-levels of abstract representations at the same time as detailed representations (e.g., Pierrehumbert, 2016; Ambridge, 2020b). When encountering speech, stored exemplars are used to interpret new utterances via activation levels dependent on a variety of factors, including similarity, frequency, and recency.

Because the same mechanisms of categorization can apply to phonological information (e.g., clusters of variants in similar phonological environments) and indexical information (e.g., clusters of variants in similar social contexts or spoken by similar people), this approach has been described as a "cognitively-realistic, integrated theory of phonological knowledge, speech production, and speech perception in which indexical knowledge is not marginalised but central" (Foulkes, 2010, p. 32). Nevertheless, G. J. Docherty and Foulkes (2014) argued that, despite its potential, certain facets of exemplar approaches remain theoretically underdeveloped, including inter-related issues of learning and acquisition, the role of the individual, and the influence of previous experience (discussed as "signal-independent" information) in modulating sensory perception via attention.

More recent approaches within sociophonetic cognition have engaged with some of these issues. In particular, a focus on the development of "attention-weighting" exemplar theoretic approaches has been propelled by research showing that frequency in experience alone does not always explain linguistic behavior (Sumner et al., 2014; Drager & Kirtley, 2016). Within these formulations of exemplar models, prior experience can include various types of linguistic exposure, such as experience gained via interaction (e.g., "authentic" speech variants), performance (e.g., exaggerated speech variants), and metalinguistic commentary (e.g., discussion about speech variants). At the same time, the presence of awareness, linked to ideologies and stereotypes, can lead to salience, while salience increases attention. In this way, awareness, while not the same as attention nor always expected to lead to attention, can be seen as cognitively related. As in the basic conception of exemplar models, phonetically-rich exemplars are encoded in memory and indexed to social information; when certain variants are more frequent, those exemplars will be larger in number. Crucially, however, exemplars and indices are also *weighted* such that more attention at the time of experience leads to stronger weights. Thus, it is possible that variants with relatively few exemplars (low exposure) that are in some way more salient (high attention) result in more robust exemplar clusters, which then more strongly influences perceptual processes. Of course, this process is highly variable and idiosyncratic, and it means that the same exposure could result in different interpretations and exemplar weighting both within and across individuals.

Under this framework, sociophonetic speech perception would proceed as follows. At the time of a new speech perception experience, listeners attend to certain salient parts of the available signal, and there is increased activation for associated (or similar) exemplars. Activation is determined by both number and weight: Those exemplars with greater number or weight will have greater activation. The perceiver then interprets incoming linguistic information based on the most highly activated forms which lead to shifted perceptual decisions, faster lexical decisions, and so forth. Those interpretations will in turn be stored and weighted based on the attention paid at the time. In this way, sociophonetic perception can be seen as a self-reinforcing cycle: Past experience, crucially involving ideological components, directs attention and shapes current subjective interpretations; these are then stored as new, weighted exemplars and influence future experiences.

Overall, this model predicts roles for ideology, stereotypes, awareness, and salience in both specific instances of perceptual processing and the development and reinforcement of sociophonetic associations over time. A more fully specified version of exemplar-theoretic models, and by extension a clearer picture of sociophonetic speech perception, requires a fuller understanding of these concepts and their impacts. As well, if these concepts are essential to accurately describing sociolinguistic perceptual processing, we must be precise in defining them; however, thus far, these terms have been used in a multitude of ways differing across sociolinguistic and cognitive psychological traditions (see Campbell-Kibler, 2016). In the following, I both discuss selected approaches to these key concepts drawn from sociolinguistic, cognitive, and social cognitive traditions and establish the approach that I take in the current work to conceptualize how they feed into perceptual representations, processing, and learning within exemplar frameworks and more broadly.

2.1.2.2 Ideology & Stereotypes

To review, I use *ideology* (Woolard & Schieffelin, 1994; Irvine & Gal, 2000) to refer broadly to ideas about social categories that exist in the world, as well as about how those social categories connect to language use, linked to moral or political attitudes about specific styles or linguistic variants (e.g. stigmatization). *Personae*, which are recognizable ideological characters representing ideas about a certain "type" of person including their speech style, are one form that ideology can take (Eckert, 2008a; D'Onofrio, 2020). Sometimes, personae may be discussed as a "stereotypical" example of a social group. I refer here to D. J. Schneider's (2005) basic definition of *stereotypes* from a social cognition perspective as "qualities perceived to be associated with particular groups or categories of people", acquired through a combination of individual experience and cultural transmission (p. 24). These generalizations are typically linked to differentiating between groups and involve folk theories about why they are true. Another way to describe stereotypes is that they are simply "semantic associations activated by social categories" (Freeman & Johnson, 2016, p. 370). In this sense, stereotypes describe components of ideology as they refer to (beliefs about) associations between social categories and features (e.g., traits, physical appearance, behavior). Personae may embody or incorporate stereotypical qualities, and in such a case could be considered a holistic "stereotype" of a certain social category. More generally, categories and associations (or stereotypes) can both exist at an unconscious level and influence behavior without active use, or they can be consciously available for people to discuss, apply, and otherwise impact behavior (Greenwald & Banaji, 1995). This aspect of consciousness brings us to the concept of awareness.

2.1.2.3 Awareness

Following Drager and Kirtley's (2016) approach, I conceptualize *awareness* in sociolinguistic representations as conscious access to (or "overt knowledge" of) social categories (e.g. Jock, Valley Girl), linguistic variants (e.g. fishin', like), and/or associations between social dimensions and linguistic variants (e.g. Canadians say "eh"). In other words, awareness acts on ideological constructs and stereotypes, describing ideas that are in some way or to some degree available in conscious cognition. The concept of awareness has long been discussed in sociolinguistic literature, though the way it is described has varied. For example, Labov's (1972) taxonomy of sociolinguistic variables proposed a tripartite system that foregrounded awareness as an important factor: (a) *indicators* that pattern by social group but without any social evaluation or stylistic variation, (b) *markers* that vary as a function of style but below awareness, and (c) *stereotypes* that vary and are above awareness, considered to be socially marked and labeled, or "salient". As well, Preston (1996) theorizes about modes of "awareness", but here the term is used to mean any knowledge, including unconscious knowledge. In this taxonomy, (a) *availability* refers to the ability to discuss a variable—thus indicating some level of overt knowledge, or awareness in Drager and Kirtley's (2016) terminology. The nature or access to sociolinguistic knowledge can also be specified for (b) degree of *accuracy*, (c) degree of *detail*, and (d) the ability to perform the style or features (*control*). Importantly, all of these aspects—including availability or conscious awareness—can exist on a continuum, resulting in an enormous range of variability for how and to what degree sociolinguistic knowledge and awareness can be present for individuals (not to mention measured by researchers). Linguistic ideologies, personae, and stereotypes are thus subject to differing degrees of conscious knowledge.

Of additional relevance is a discussion of awareness not just in terms of representations but in terms of processing and research task design, as in "implicit tasks" (see De Houwer, 2006). In the context of a particular task, implicit processing has been used to refer to participants processing or responding without (a) being aware of the target concept (e.g., a particular sociolinguistic variable or association, as discussed above), (b) being aware of the fact that the target concept is *being measured* in the task, and (c) the ability to control responses, regardless of conscious awareness. Under certain approaches, implicitness is the same concept as automaticity, where lack of awareness (i.e., outside of consciousness) and control (i.e., cannot be strategically changed or stopped) are two aspects of automatic processing (Bargh, 1992; De Houwer, 2006; De Houwer & Moors, 2010). Thus, processing without awareness or controlled responses can be used to describe a form of automatic processing. These distinctions help to clarify that a lack of awareness within (sociophonetic) perceptual processing could refer to any of the following scenarios: (a) a listener not having conscious access to a particular ideological construct or association in general, (b) a listener automatically processing and responding to speech without conscious access in the moment (regardless of their general state of overt knowledge), and (c) a listener gaining consciousness in the moment but still without control to apply that knowledge, thus responding automatically in that sense. It is thus important to specify whether we are discussing awareness as applied to representation or processing, as both high or low representational awareness may be paired with conscious or unconscious processing and these different combinations may have different consequences.

If the broad definition of awareness entails some form of conscious cognition, it is also relevant to clarify our approach to consciousness. Under one view, conscious access can be described as when "a piece of information ... becomes broadly available for multiple processes including action planning, voluntary redirection of attention, memory, evaluation, and verbal or non-verbal report" (Dehaene et al., 2003). That is, unlike unconscious content which is confined to localized perceptual processes, conscious content crosses a threshold to become broadcast to various cognitive processes and therefore is "globally available" (see Dehaene et al., 2003; Mashour et al., 2020). During perception, consciousness occurs only if bottom-up stimulus strength is sufficiently strong to exceed the activation threshold *and* top-down attention is present (Dehaene, 2011); otherwise, processing remains subliminal (in the absence of sufficiently strong stimulus strength) or preconscious (in the absence of attention). Early stages of perceptual processing (e.g., the first 200 ms) are always unconscious, but if consciousness is reached, later processing involves a stronger and longer-lasting neural response. Taking this approach, being aware of a sociophonetic variable or association means those representations are in a state of global availability. In addition, this view of consciousness where attention is a conditioning variable aligns well with attention-weighting exemplar models, whereby attention is a key selective mechanism leading to more robust encoding of certain stimuli or experiences. Next, I consider the closely-tied concept of salience, which can also be linked to attention.

2.1.2.4 Salience

Within sociolinguistic traditions, *salience*—similar to awareness—has been long discussed, but has also been variably defined and approached. In earlier dialectology and language change work, salience was often characterized by criteria or factors contributing to it (see review in MacLeod, 2015); some of these alluded to awareness in how speakers perceive differences between varieties, such as stigmatization (Trudgill, 1986) and overt stereotyping or mimicking (Auer et al., 1998). The specific role of awareness as a factor in salience is also recognized in cases where they are essentially equated, as in one definition of salience being "degree to which speakers are aware of some linguistic feature" (Hickey, 2000, p. 57). Later work narrowed down the definition of salience to the ability to stand out, as in a property or characteristic of being prominent or noticeable (Kerswill & Williams, 2002). Separately, a more applied approach has been taken to define sociolinguistic salience as degree of social meaningfulness or sociolinguistic association, regardless of conscious awareness (e.g. MacLeod, 2015; M. M. Jensen, 2016); in some approaches, community agreement on social meaningfulness is an important component (Llamas et al., 2016). These two main views in sociolinguistics are summarized and termed by Rácz (2013) as (a) *cognitive salience*, the property of being prominent or noticeable leading to unexpectedness or surprisal, and (b) *social salience*, indicating a marker of social indexation (originally sourced from cognitively salient features).

Cognitive perspectives on salience in processing (e.g. Schmid & Günther, 2016; Zarcone et al., 2016; Günther et al., 2017) offer additional insights with a slightly different angle and usage of terminology. Günther et al. (2017), for example, propose five dimensions of salience relevant to psychological and linguistic research. One of these is *domain*, where salience can be defined in the perceptual domain as "the distinctiveness of an item relative to its surrounding items" or in the cognitive domain as "the accessibility of a stimulus in long-term memory" (p. 302); in the former, salience attracts attention, while in the latter, salience focuses attention. Another dimension is *loci* such that salience can be defined from the perspective of individuals or collectively at the community level. In addition, there can be bottom-up (external) *sources* of salience which are driven by stimulus-inherent features, and top-down (internal) sources which are based on expectations. Focusing on different instantiations of linguistic salience, Schmid and Günther (2016) narrow down two main elements of top-down salience: sources of expectations and mechanisms of salience. Within this framework, salience "emerges from a comparison between an incoming linguistic cue and expectations" where expectations can come (mainly) from (a) general long-term memory storage or (b) current (linguistic, situational, and social) context while the mechanism of salience can be (a) confirmation of expectations or (b) violation of expectations (Schmid & Günther, 2016, p. 4). This results in the possibility of both context-free and context-bound salience due to either entrenchment or surprisal.

Taking all these perspectives together helps to characterize salience as relevant to sociolinguistic awareness and perceptual processing. Both the existence of social indexation and awareness of those indices are factors that contribute to certain expectations in certain contexts. When those expectations activate due to social context (e.g., inferred speaker identity) and an incoming linguistic cue either clearly matches or clearly mismatches those expectations, that cue becomes salient. This context-based salience can be viewed both through the lens of particular individuals, or aggregated across a community. The next step then is to integrate these insights about salience into an understanding of how sociophonetic associations play into processing and learning.

In one proposed framework, Günther et al. (2017) suggest that salience and memory form a feedback loop (see schematization in Figure 2.1). If we take the

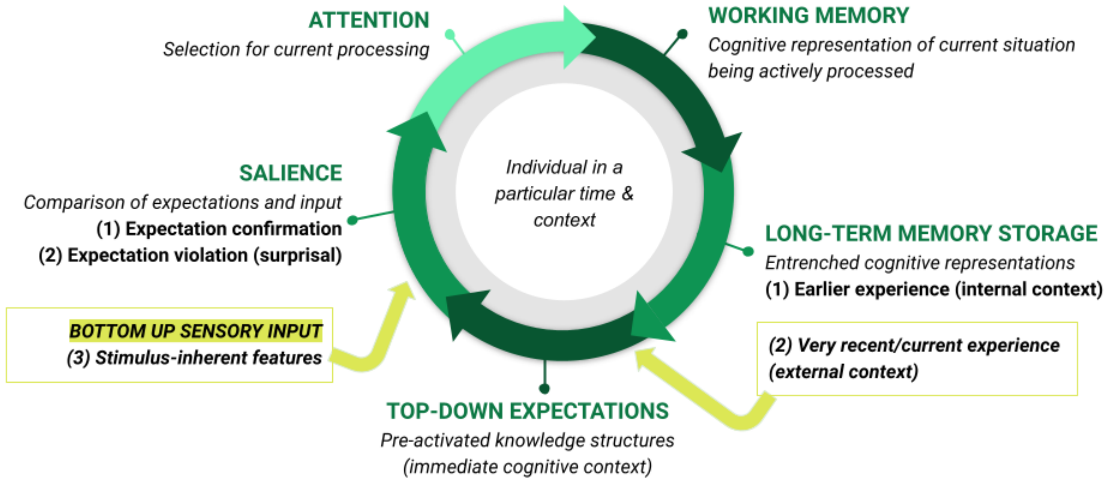

Figure 2.1: Schematic diagram representing a feedback loop between salience and memory (based on Schmid & Günther, 2016; Günther et al., 2017).

starting point of any particular perceptual experience to be long-term memory storage, representing cognitive representations from past experience, that together with the current external context may lead to certain top-down expectations. Those expectations consist of pre-activated knowledge or representations that are compared to the incoming linguistic input, and these either match or mismatch, leading to salience; alternatively, stimulus-inherent features from the input result in salience. Either way, these aspects of the input draw attention, and thus are selected for processing in working memory, and then are likely to become entrenched in long-term memory, therefore affecting any further perceptual experience.

This outline of a feedback loop as applied in general terms to perceptual processing is highly similar to that proposed within attention-weighting exemplar-theoretic frameworks as specified for sociophonetic perception: Earlier experiences and current experiences serve as a self-reinforcing cycle mediated by salience and attention. Indeed, these can be seen as different framings of the same proposal, although the more general salience feedback loop does not directly consider stereotypes and awareness. If we apply it to sociolinguistic perception, ideologies, stereotypes, and personae that link social categories to linguistic cues comprise the relevant cognitive representations stored in memory. Where awareness may come into play is by strengthening the cognitive processing of any particular socio-perceptual experience that crosses the threshold into consciousness within working memory, and therefore encouraging strengthened long-term memory entrenchment. Another option would involve awareness being acquired outside of direct perceptual experiences (e.g., via metacommentary), which also results in relevant representations being stored and/or activated (and therefore strengthened) in long-term memory (Drager & Kirtley, 2016). Both scenarios should lead to increased expectations (i.e., pre-activations) resulting in greater salience and attention with influences on memory, which contribute to increased expectations in the next iteration. Altogether, the salience-memory feedback loop provides a useful schematization that is compatible with exemplar-based models for how both sociophonetic processing occurs, and how sociophonetic associations are learned and reinforced through later processing.

In summary, if we view perceptual experiences as a feedback loop of expectations, attention, and memory updating, then awareness serves to strengthen encoding of sociolinguistic associations (including in the form of stereotypes or personae) via accumulated salience over time. In other words, this approach predicts a role for awareness in both influencing in-the-moment perceptual outcomes and acquisition of sociolinguistic variables over time. However, empirical evidence for this model remains limited, and some aspects would benefit from further development. I next expand on the background provided in the last two sections, highlighting areas for further empirical study that could help to support and develop attention-weighted exemplar models of sociophonetic speech perception.

2.1.3 Directions for Further Research

Three areas that intersect with exemplar approaches to sociophonetic perception and that could benefit from further clarification and empirical evidence include: what social categories are available to listeners, how awareness and salience mediate perception, and how socioindexical associations are learned. For each, I discuss the theoretical relevance, relevant literature, and gaps in our understanding, as well as consider potential paths forward.

2.1.3.1 Nature of Social Representations

Social dimensions that are learned, cognitively represented, and linked to linguistic styles are much more complex than a set of broad macrosocial category labels (e.g., Campbell-Kibler, 2007; Eckert, 2008a; D'Onofrio, 2020; Drager et al., 2021). This acknowledgment problematizes overly simplistic approaches to social meaning that have traditionally shaped interpretation of sociophonetic results, and allows for more direct and cognitively realistic accounts of the mechanisms driving sociolinguistic behavior. Exemplar models can accommodate different forms of social information, but they do not dictate details of the cognitive representations and their relationships. For any particular context, what social constructs do listeners actually access, given what we know about linguistic ideologies and experience? How are ideological categories cognitively structured relative to each other, and when are each evoked during perception? How might different types of social categories differentially influence perceptual processing? These questions help to elucidate details of sociolinguistic cognition that inform our understanding about speech perception and motivations behind language change.

Perspectives from social meaning approaches to language variation and change provide the insight that while broad social categories appear to modulate perceptual responses (and behavior in production) in the aggregate, there are multiple possibilities for what is happening at the level of individual cognition. In this view, linguistic styles and the personae that exhibit certain styles comprise the main ideological foundations at the individual interactional level, directly influencing interpretation (Eckert, 2008a); only when scaling up to the aggregate level do we see patterning based on broad macrosocial categories, *indirectly* arising from style and personae-based behavior. At the same time, conceptions of macrosocial structure exist ideologically (e.g., institutionally) and do inform microsocial representations and interactions (see D'Onofrio, 2020). Further, a macrosocial label may come to be (most) strongly linked to a particular "prototypical" persona, ideologically seen as characteristic of that demographic group, rather than being inclusive of the broad category. For example, rather than a broadly inclusive local identity, classic "New Yorker" and "Chicagoan" identity may be most associated with certain local white personae and their characteristics, which may motivate other locals to orient away from those linguistic styles (Becker, 2014; D'Onofrio & Benheim, 2020). This altogether complicates our understanding of individual level social representations and the motivating expectations that impact perceptual processing behavior.

Evidence of the microsocial level of representation held by listeners is found in studies showing that, along with broad demographics, a wide range of personae, personality attributes, behaviors, physical appearance, stances, and more can be interpreted from certain linguistic styles in short samples of speech (Campbell-Kibler, 2007; Drager et al., 2021). For example, Drager et al. (2021) demonstrate the many possibilities for narrower, intersectional, and highly localized categories of people in the Hawaiian context and how these and other social judgments are both complexly and relatively consistently modulated by a voice's pitch across listeners.

MacFarlane and Stuart-Smith (2012) similarly showed that linguistic styles and specific linguistic features are cognitively linked to highly localized stereotypes—in this case, two types of Glasgow men ("Glasgow Uni" and "Glaswegian") that are schematically associated with particular brand logos. In a matched guise speaker identification study, listeners who were also Glasgow men reliably differentiated speech samples based on the two personae, and showed this effect to different degrees for four phonetic variables of varying local salience. Moreover, in examinations of socially primed perception, D'Onofrio (2018b) show that attributions of personae such as "Valley Girl" and "business professional" to a speaker modulate listeners' perceptual phoneme boundaries for the TRAP vowel, indicating that listeners use social constructs at this level to form perceptual expectations of fine phonetic distinctions. D'Onofrio (2019) also demonstrates that images of the same East Asian man presented with two different visual styles elicited different social evaluations and perceptual recall performance. This poses a problem for the interpretation and generalization of previous studies using the broad category of "East Asian" race/ethnicity represented by a single photo, which were discussed in the context of generalized "non-native" stereotypes of "Asian" individuals (e.g., Rubin, 1992; M. Babel & Russell, 2015; McGowan, 2015). Exactly what ideological constructs were listeners in those studies tapping into to form expectations and inform their perceptual processing? What other personae or styles of individuals may evoke different processing outcomes?

For a finer-grained understanding of the cognitive representations and processes driving individual and group-level perceptual patterns, it is imperative that sociophonetic perceptual studies carefully consider listeners' ideologies about categories of people that underlie macrosocial generalizations and their associated linguistic styles. Here, awareness and salience may play a role in strengthening the ideological connection of different linguistic variables with particular styles (see MacFarlane & Stuart-Smith, 2012). Particular attention could be paid to how a macrosocial label or prototypical example of a macrosocial group mediates expectations relative to more narrowly- and/or intersectionally-defined personae subsumed under that group. Generally, this approach may provide more ecological validity: Empirical evidence of perceptual outcomes in the face of more contextualized social categories, compared to contextless macrosocial identities, may be more generalizable to speech perception in the social world. Overall, assessing a fuller picture of the ideological background listeners are bringing to a perceptual task could help future research to make more informed interpretations about the data.

2.1.3.2 Awareness and salience of sociophonetic representations

Another aspect to be explored further is the part that metalinguistic awareness (and the closely-linked idea of "social salience") plays in speech perception. As discussed in depth already, recent exemplar-based approaches link awareness to salience then attention as a factor modulating encoding and later pre-activation of exemplars to influence perceptual processing outcomes; however, open questions remain about the exact nature of awareness and its interaction with other mechanisms and processes (Drager & Kirtley, 2016). For example, awareness itself—the conscious access to sociophonetic information—is considered to be able to, but does not always, contribute to increased salience. While awareness is often seen as a key indicator of salience in sociolinguistic approaches, what is the actual extent of connection between these separate-but-related constructs, in general or in different contexts? Framed another way, what are their individual contributions to sociophonetic representations and processing? As well, awareness can be acquired via different routes, such as through exposure and attention to items that are in some other way salient, through metalinguistic discussion without exposure, or a combination of the two. Would awareness gained through different types of experience result in quantitatively or qualitatively different outcomes for representations and processing? Because salience and awareness are such central concepts in language variation and change, improved description of their function within sociocognitive processes should have widespread benefits.

While salience has commonly been called upon in interpreting results, direct investigation of its influence has been less frequent until recently; in these, the link to awareness is often implied or directly tied to salience. Regardless, the results have been inconclusive in terms of the prediction that higher salience should result in greater perceptual influence. Some studies consider salience as properties of sociolinguistic variables. In one case, Lawrence (2015) reported unexpected null results for a social priming perceptual study involving highly salient (overtly stereotyped) variants in Northern and Southern British English. On the other hand, Juskan (2018) compared four Standard Southern British English (SSBE) variables with varying levels of local salience, reporting social priming only for the higher salience variables, as predicted. In the case of a low-salience gender-linked sound change, Alderton (2020) found no overall priming effect on /u/-fronting in SSBE which is consistent with hypotheses, though a small effect for listeners who were men *was* found. When looking at salience across individuals, Chang (2017) did not find priming effects based on individual salience in a perceptual study of Taiwan and Beijing Mandarin regional variation. At the same time, I found in previous work that individual-level degree of awareness for the stereotyped difference between Canadian and Michigan /aʊ/ did seem to mediate socially-primed perceptual responses, at least when mediated by degree of cognitive empathy, representing one aspect of individual sociocogntive processing style (L. S. P. Cheng, 2021).

Potentially contributing to the inconsistencies in findings is that the method of assessing sociolinguistic salience and/or awareness has been highly variable. Some studies have assigned a level of salience to particular linguistic variables based on existing knowledge of local overt stereotypes (Lawrence, 2015), or language-internal and -external criteria from prior literature, including whether the variable elicited metalinguistic commentary (Alderton, 2020). Juskan (2018) assessed traditional Labovian criteria corresponding to the *indicator-marker-stereotype* distinction—namely, social stratification, hypercorrection, style shifting, and overt comments or evaluation—comparing across multiple variables; in this sense, salience equated to awareness of a sociolinguistic variable. Chang (2017) attempted to measure salience individually with a categorical measure: Participants were asked in an open-ended question to report the most salient difference between Taiwan and Beijing varieties of Mandarin, and participants were grouped based on the difference they named. For the study reported in L. S. P. Cheng (2021), I assessed gradient degree of individual awareness via a composite variable encompassing multiple questionnaire items designed to assess degree of stereotype knowledge. The questions included self-rated familiarity with the stereotype, self-rated frequency of exposure to the stereotype, and scores on an assessment where participants were asked to select words that they believed would be pronounced differently (presumed to reflect one form of stereotype knowledge).

Overall, the evidence in favor of salience influencing perception is inconsistent, but the methods used to infer awareness or salience have been equally inconsistent. Also, as few studies are interested in the *causes* of salience, rather focusing on its *effects* (e.g., Juskan, 2018), how salience or awareness was acquired and how awareness as a factor contributes to salience have not typically been probed. Even if they are expected to correlate to some extent, the conflation of social salience and awareness limits our understanding of how they interact cognitively and their independent contributions to perceptual outcomes.

Future studies may consider disentangling these constructs and measuring them separately. Improvement in direct assessments of salience and awareness in addition to some extent of standardization across studies would help to more carefully test salience-based hypotheses, as well as flesh out details about how awareness works socio-cognitively. As well, a move towards direct and gradient assessments of awareness and salience would be useful. Relying on previous reports of a stereotyped phenomenon can be misleading as awareness and salience can change over time, differ across communities, and of course vary by individual. At a more basic level, overt commentary signaling the *presence* of metalinguistic stereotype awareness does not allow insight into either degree of awareness (L. S. P. Cheng, 2021) or salience (Llamas et al., 2016), and may lead to inflated estimates of the strength of sociolinguistic associations that exist. For example, a large group of Michiganders may be able to report that they believe Canadians and Michiganders pronounce the word "about" differently if prompted (or even unprompted in open-ended responses), but the extent to which this overt stereotype is familiar and cognitively accessible can differ gradiently. One possibility to directly assess social salience for linguistic variables is proposed by Llamas et al. (2016): In the Social Category Association Test (SCAT), the speed of categorizing a linguistic variant to a social category together with group-level agreement in categorization is taken to represent the degree of social meaningfulness, encompassing both conscious and unconscious associations. They find that certain overt linguistic stereotypes of Scottish English did not demonstrate strong associations based on the SCAT relative to other variables, arguing in support of direct measurement of salience.

While considering awareness and salience at the individual level is perhaps most direct, many factors may make this impractical to study "in the wild", including a lack of validated and reliable measures which may make potentially small effect sizes even more difficult to detect. Broadly speaking, however, attention-weighting models predict effects of (social) salience on perception whether it varies across particular linguistic variables, lexical items, speech communities, or individuals. A wider variety of studies with different loci of salience and awareness can help to generally test the hypothesis. For example, comparing the same task and variable across communities who differ on overt stereotyping (e.g., Australians versus New Zealanders; see Hay & Drager, 2010; Walker et al., 2019) or lexical items that are more or less associated with overt stereotypes (e.g., "fish and chips" in Australian versus New Zealand English; see Hay, Nolan, et al., 2006) could help to triangulate the influence of stereotype awareness and/or salience on perception. Another option besides quasi-experimental studies that rely on reliably measuring existing awareness and salience tied to variables and individuals is to use experimental manipulation: Experimental groups exposed to conditions encouraging awareness of stereotypes would be predicted to differ in salience and perceptual outcomes from those who were not exposed.

2.1.3.3 Learning of sociophonetic associations

Finally, the acquisition of sociophonetic variation is a topic that has historically been understudied, though much more work is emerging in recent years (Foulkes, 2010; G. J. Docherty & Foulkes, 2014; De Vogelaer et al., 2017). While exemplar theory (and other usage-based approaches) provide a general framework for how development and learning can be accounted for, the details of how it unfolds over time and how exemplar storage is updated in memory requires further refinement. For example, how much exposure is needed to learn a sociolinguistic association? Beyond exposure, what factors mediate the aspects of an experience that is learned (i.e., attended to, encoded, and stored)? Specifically, what is the role of awareness and salience in learning socioindexical or contextual meaning, with and without further exposure? While theorized to influence learning, few studies have directly assessed awareness and/or salience. Answers to these questions would help to explain who learns what associations and when, relevant to various areas of study including early language development, second language acquisition, and language change.

A growing number of studies have tested various aspects of learning socioindexical associations in experimental and artificial language experiments, particularly highlighting the role of salience in promoting learning of social meanings as well as the prominence of individual variation. Using an auditory task, G. J. Docherty et al. (2013) found evidence that adult British English listeners were able to learn associations between certain consonant or vowel variables and arbitrary social category labels (e.g., "tribe1" versus "tribe2"), but that this varied by individual, linguistic variable, and categoricity of the variants. Importantly, the most clear and consistent learning was found for the locally sociolinguistically salient variable of [t] and [ʔ]. The majority of other studies utilize text-based presentation of language and consider the salience of social rather than linguistic properties but corroborate these general findings.

In a series of studies on contextual morphological variability, Rácz and colleagues test learning based on preexisting social categories with an artificial language. They find that not only does a socially salient speaker property, namely gender, result in more association learning than a non-salient, incidental speaker property (spatial orientation; Rácz et al., 2017), certain social categories appear to be less socially salient (ethnicity and age, as compared to gender) and result in less evidence of learning with the same amount of training (Rácz et al., 2020, see also Li & Roberts, 2021). Notably, their results distributions were always bimodal rather than gradient: Only some individuals learned the associations ("good learners") while others did not show learning at all, but the number of "good learners" increased with a more salient social cue and additional training. In further support of the role of salience, Sneller and Roberts (2018) reported learning and use of a particular association only when it was *socially relevant* to the task—or in other words, socially salient—while Lai et al. (2020) demonstrated that another form of salience, driven by unexpectedness rather than social properties, facilitates the learning of sociolinguistic variation.

This line of research would benefit from a greater variety of studies testing who learns which types of sociolinguistic variables associated with which types of social categories. Thus far, the pool of evidence is still limited and contained to relatively artificial settings. Additionally, the learning outcomes measures have either assessed overt social categorization or productive usage rather than perceptual behavior. Further research will be needed to confirm whether these results extend across sociolinguistic contexts, particularly a more ecologically valid test case (e.g., an existing sociolinguistic category and potential linguistic association) and into the domain of auditory sound-based linguistic variation. Since salience appears to be a key factor, if metalinguistic awareness increases salience, then the presence of awareness should also be linked to an increase in the speed of learning and number of individuals who learn an association. Again, however, the direct connection of awareness to acquired social meaning and use in processing requires empirical evidence.

In all, several areas of research concerning sociophonetic representations and processing reveal gaps in our understanding, particularly as applied to more ecologically valid sociolinguistic contexts. In the next section, I provide a big picture view of the specific context that will be examined in this work: Asian American/Canadian speech.

2.2 The Asian American/Canadian Context

The sociolinguistic context of Asian Americans and Canadians is situated within ideologies of race and ethnicity within broader North American society.

Generally, *ethnicity* is considered to refer to socially constructed categories based on cultural difference while *race* refers to a social construct based on physical appearance; however, racial ideology conflates cultural aspects with physical characteristics in an essentializing framework that positions both biological and behavioral differences as linked, innate, and immutable (Smedley & Smedley, 2005). One consequence is that the line between ideologies of race and ethnicity is sometimes unclear, especially so because these constructs are flexible, changing, and dependent on context (e.g., census differences in race and ethnicity questions across place and time, Stevens et al., 2015). Still, racial and ethnic ideology have meaningful societal and psychosocial consequences for individuals; these are tied both to external forces, such as how outsiders' categorize people as members of a social group, and internal forces, such as how people self-identify with a social group (Renn, 2012).

These issues are evident within Asian American/Canadian experiences. "Asian American" as a term was first coined by activists in the 1960s as part of a political movement encouraging pan-ethnic solidarity (see Kibria, 1998; Okamoto & Mora, 2014; J. Lee, 2019). Due to lobbying in the 1980s and 1990s, "Asian" is currently used as an official census racial category in the US (J. Lee, 2019), and thus institutionally encoded. In line with this, "Asian American" and "Asian Canadian" can be defined as a comprehensive demographic grouping of individuals of Asian origin in the US and Canada, including ethnic groups labeled based on geographic origin as East Asian (e.g., Chinese, Japanese, Korean), Southeast Asian (e.g., Filipino, Vietnamese, Cambodian), South Asian (e.g., Indian, Pakistani), and sometimes West Asian (e.g., Iranian or Persian, Armenian, Lebanese). As a racial category, such a grouping is often seen by outsiders as a homogenous, monolithic group with inherent shared physical and cultural characteristics (Reyes & Lo, 2009). As a broad pan-ethnic category, it can be seen as representing diverse groups and individuals tied together by a shared experience and history of racialization (Kibria, 1998). Ethnic categories would refer to distinct cultural groups underlying the broad category of "Asian American/Canadian".

In practice, however, distinguishing the concepts of race and ethnicity relative to the ideological construct of "Asian American/Canadian" is not so straightforward, and the use of the term as an umbrella category for such diverse demographics can be problematic as well. First, the boundaries of who "counts" as Asian American is variable along dimensions of generation, socioeconomic class, ethnic background, and more. Importantly, tied in part to North American immigration history, "Asian American" identity tends to be interpreted as, or have the strongest association with, East Asian identity. In other words, both for those who identify as "Asian American" and those who do not, East Asians such as Chinese, Japanese, and Korean are most likely racialized or accepted as "Asian American", followed by Southeast Asians and finally South Asians (Park, 2008; J. Lee & Ramakrishnan, 2020). Second, the adoption of the pan-ethnic label is not uniform across those who may be racialized as such. For some people, particularly those raised in North America, "Asian American/Canadian" may be used as a social group and self-identification label representing perceived shared culture, values, and second-generation cohort experiences (Park, 2008; J. Lee, 2019). Other individuals do not accept this pan-ethnic label as part of their identity, and uptake can vary by ethnic groups (e.g., Lien et al., 2003). Third, there are multiple aspects of identity that can coexist either in internal self-ascription or external categorization. For example, many individuals appear to primarily identify with a specific ethnicity (e.g., Vietnamese or Vietnamese American), but also accept a broad racial or pan-ethnic identity (e.g., Asian American; Lien et al., 2003). Other possible identities include intermediate pan-ethnic categories (e.g., "Southeast Asian", "Chinese", "Desi") as well as intersectional racial or ethnic identities (e.g., "Asian woman"). An individual may highlight one identity more than another in different contexts, or embody multiple at once. Similarly, in perception, a perceiver may attune to or make judgements based on different racial or ethnic categories at different times, or possibly multiple at once; exactly what category a perceiver is attending to (e.g., Asian, East Asian, Korean) at any given time is difficult to know.

My approach to race, ethnicity, and the Asian American/Canadian construct is informed by these insights. In this work, I will commonly use combined terminology such as "race/ethnicity", "racioethnic", or "ethnoracial", acknowledging that the distinction between race and ethnicity can be rather murky. In addition, given that the main focus of this dissertation will be on how ideas about social groups and their speech influence linguistic processing, "Asian American/Canadian" (abbreviated to AAC) will generally refer to a category based on typical ideological interpretations—whether referring to how speakers use it for identity work, or how listeners use it for interpretative work. In other words, "Asian American" and "Asian American speech" in this work will most likely refer to East and Southeast Asians—those who are most ideologically associated with "Asian American"; it may also most likely refer to second-generation individuals and those of a particular socioeconomic status, and any other assumptions tied to the ideological construct. I use the pan-ethnic term while recognizing that this research does not speak to, and will not generalize to, the experience of all individuals who may self-identify as AAC or fit under the broad demographic category, but also recognizing that circulating ideologies about AACs and their speech *do* exist and could impact the experiences of many.

To take one example, AACs are commonly stereotyped with foreign status (e.g., "forever foreigner", Tuan, 1998) and "accented" or "non-native" speech (e.g., "Mock Asian", Chun, 2004), which can be negatively evaluated on a range of traits (e.g., Cargile, 1997; Lindemann, 2003; Hosoda et al., 2007; Bauman, 2013). Accordingly, some research indicates that the stereotyped expectations of "accented" speech from an individual racialized as "Asian" affects how linguistic processing and memory proceeds (e.g., Rubin, 1992; M. Babel & Russell, 2015; McGowan, 2015). Beyond this, AACs who are recognized as local may also be linked to particular speech styles (e.g., Hanna, 1997; Newman & Wu, 2011) and stereotypes (e.g., Pyke & Dang, 2003; M. Jeon, 2007), which may or may not be linked to the idea of foreignness. However, these ideologies, including AAC-associated linguistic features and personae, as well as their impacts on socio-perceptual representations and processing are only sparsely represented in the literature. More nuanced understanding of these dynamics would shed light on the social and linguistic experiences of AAC individuals, an overall understudied topic (Reyes & Lo, 2009). Detailed background relevant to AAC speech production, perception, and ideology will be reviewed within Chapters III and IV. With this in mind, I move forward to summarize the research goals for this dissertation.

2.3 Summary of Research Goals

Sociophonetic research has highlighted the connection between social and linguistic information, which can be conceptualized as ideological associations or stereotypes, and its impact on perception. Attention-weighted exemplar models can be used to conceptualize how ideology may influence perceptual representations and processing via awareness, salience, and attention. However, development of our understanding of sociophonetic perception could benefit from increased empirical study and approaches that move beyond macrosocial categories, directly assess and account for awareness and salience, and address learning of sociophonetic associations.

As such, one main goal of this dissertation is to explore some of these topics further, examining how personae are linked to speech styles and mediate perceptual processes, as well as investigating the links between stereotype awareness and perceptual behavior. In particular, I do this within the context of AAC linguistic ideologies and stereotypes. As prior sociophonetic research on AACs has been relatively sparse, the other major goal is to add to existing literature on AAC speech, including production, perception, and ideologies, to more fully characterize how AACs are cognitively represented and socially perceived by others.

Firstly, I seek to establish ideologically-relevant features of AAC speech, asking what phonetic features are potentially associated with AAC identities, and to what degree of awareness. In addition, I explore social personae that may exist under the broader macrosocial AAC category, and how they are linked to speech styles and linguistic variables. Then, I target the role of microsocial categories in sociophonetic representations, testing in a novel context how personae may mediate socio-perceptual processing of specific linguistic variables. At the same time, I directly assess degree of awareness and its role in modulating strength of socio-perceptual outcomes, considering these both at the level of linguistic variables and individuals. On the whole, this dissertation aims to contribute a more in-depth understanding of how AACs and their speech are represented both ideologically and cognitively. At a more generalized level, it aims to expand our understanding of sociophonetic cognition, providing new empirical evidence to fine-tune current models of sociophonetic representations.

CHAPTER III

Exploring Production of AAC Speech

3.1 Introduction

With the goal of investigating AAC speech from the perspective of listeners, I begin by examining one possible source of sociophonetic associations, namely speech production. This chapter describes an exploratory production study that examines three phonetic features potentially linked to the social meaning of AAC identity. Previous research has connected AACs to backed and monophthongal /oʊ/ as well as speech rhythm with less durational variability (e.g., Bauman, 2016). Through both descriptive and clustering analyses, I survey the variability in production patterns of these features across a sample of self-identified Asian American speakers from a variety of backgrounds, examining both group-level and individual-level patterns. The data come from a corpus of YouTube videos, representing a sample of AAC speech that is also a realistic source of direct exposure for many people. These results provide empirical evidence from a novel, naturalistic data source that contributes to clarification of which linguistic features may be, or have potential to be, sociophonetic markers of AAC identity tied to awareness and/or social personae.

3.2 Background

A small number of perceptual studies find that some listeners can identify certain speakers as "Asian" (e.g., Hanna, 1997; Newman & Wu, 2011; A. Cheng & Cho, 2021) or as a particular Asian ethnicity (e.g., P. Wong & Babel, 2017; Nagy et al., 2020) at rates above chance. This indicates that listeners have access to some form of knowledge about what AAC speech is, though the exact nature of this knowledge is unclear. For example, this could suggest that there are enregistered varieties of AAC English, or that there are particular linguistic features indexed to either a general AAC identity or a more specific ethnic identity (e.g., Vietnamese).

Regardless, features that are associated with AACs in listeners' representations may have various sources, including interactive (e.g., listening to AACs speak), performative (e.g., imitation of AAC speech), or metalinguistic (e.g., discussion of how AACs speak) experience. Recall that the most prevalent ideological associations of the label "Asian American/Canadian" involve East Asians followed by Southeast Asians (Park, 2008; J. Lee & Ramakrishnan, 2020), so the AAC identity and speech I am exploring will tend towards representing those ethnic groups represented in mainstream ideologies, rather than all groups who may self-identify as Asian American. In the following, I summarize briefly what the literature suggests about the existence of AAC ethnic speech styles, in production and in ideology. Then, I consider what specific phonetic features may be ideologically linked to AAC identity, sourced from evidence of AAC speech production.

3.2.1 AAC Ethnic Varieties

Within production, the notion of a distinctive AAC ethnolect (e.g., "Asian American English", similar to, for example, African American Englishes or Chicano Englishes) is considered to lack evidence (Reyes & Lo, 2009). This is unsurprising given the vast heterogeneity of cultural and linguistic heritage subsumed under the AAC umbrella. Nevertheless, the lack of evidence thus far does not completely rule out the possibility that pan-ethnic speech styles have developed in specific contexts, particularly in locations with highly diverse and dense AAC populations (e.g., urban metropolises). Similarly, there may exist recognizable ethnic-specific varieties of English in certain communities, such as ethnic enclaves. Indeed, in studies of specific communities, AAC speech patterns have been found to diverge from other local groups, interpreted to signal local ethnic identity and/or internal group belonging (e.g., A. W.-M. Wong & Hall-Lew, 2014; M. Zheng, 2018; Sheydaei Baghdadeh, 2021). As such, listeners may have stored representations of specific ethnic speech styles and/or features of the (local) communities they have contact with (P. Wong & Babel, 2017; Nagy et al., 2020).

However, other evidence suggests that listeners do not always have access to specific ethnic varieties, though they may have access to a broader “Asian” category. Newman and Wu (2011) reported that listeners in New York City were not able to distinguish between Chinese and Korean American speakers, though on the whole they were able to identify them as Asian American. In their sociophonetic analysis of the same speakers’ productions, Newman and Wu report a few phonetic cues that were shared by some of the Chinese and Korean American individuals, though none of the speakers show all the cues, nor were they fully consistent even within ethnicity. They suggest that, rather than identifying ethnicity based on any single prominent feature, listeners may be relying on a constellation of weakly represented phonetic features that, when aggregated, are associated with AAC identity. If so, what features are included in that constellation?

3.2.2 Potential Features of AAC Speech

Although the overall size of the literature is small and no established set of features have been empirically associated with AAC identities, a number of phonetic variables have been examined in production with regards to various AAC speech communities. One of the more commonly examined variables is the mid back vowel /oʊ/ in terms of both vowel quality and dynamism. According to Chun (2004), both /oʊ/-backing and monophthongization can be found in realizations of "Mock Asian" English, suggesting that they may be ideologically linked to Asian identities. In one sociolinguistic study, Bauman (2016) recorded a group of Asian American women from various ethnic backgrounds, recruited from an Asian-interest sorority in New Jersey. She found that AAC sorority members, as compared to non-sorority members, produced more backed and monophthongal /oʊ/, contrasting with the local norms in the Mid-Atlantic region and seeming to serve as an identity marker linked to the Asian American sorority.

While Bauman's study involved a rather specific context, the same features have been reported in the speech of AAC individuals in a variety of other contexts, including /oʊ/-backing for Korean Americans in California (A. Cheng et al., 2016; A. Cheng, 2020), Texas (L. Jeon, 2017), and Georgia (D.-E. Kim, 2021); note that in these locations, /oʊ/ is undergoing fronting as part of regional sound changes, which makes /oʊ/-backing even more of a departure from the mainstream. In addition, ethnic orientation or stronger identification with Korean background has been linked to more /oʊ/-backing (L. Jeon, 2017; A. Cheng, 2020; D.-E. Kim, 2021), and women have been noted to produce more /oʊ/-backing as well (Korean Americans in A. Cheng, 2020; Japanese Americans in D'Onofrio & van Hofwegen, 2020).

On the other hand, Chinese Americans have been reported not to markedly differ from Californian European Americans, fronting /oʊ/ in alignment with local vowel shifts (Hall-Lew, 2009; A. Cheng et al., 2016). Further, evidence for /oʊ/-monophthongization as an ethnolinguistic feature has so far been more limited. Still, D'Onofrio and van Hofwegen (2020) found more monophthongization in the speech of ethnic Japanese speakers in California than European American comparisons, while D.-E. Kim (2021) reports a similar but non-significant trend for Korean Americans. Altogether, the literature indicates that /oʊ/-backing and monophthongization may be used (or recognized as) an AAC ethnic identity marker, though it potentially applies only to certain ethnic groups with certain types of language contact (e.g. Korean, Japanese).

Prosodic rhythm has been another common variable of interest within AAC English, present in ideologies about "sounding Asian". It is frequently referenced in metalinguistic discussion of "sounding Asian", such as descriptions of "choppy", "jerky", or "staccato" speech (Hanna, 1997; Bauman, 2016), as well as manipulated in "Mock Asian" performances, described by Chun (2004) as "syllable-timed rhythm". One of the phonetic correlates has been suggested to be lower vocalic durational variability: the production of more uniform duration across syllables (e.g., lack of vowel reduction in unstressed syllables) rather than higher variability across syllable duration (e.g., with long stressed syllables and short unstressed syllables). Lower vowel variability has also been reported as a feature of Latin/Hispanic English in comparison to the higher variability of European and African American speakers in North Carolina (Thomas & Carter, 2006).

However, evidence linking durational variability to AACs appears rather inconclusive, even within ethnicity. In support, Bauman (2016) reports that Asian American women with a variety of ethnic backgrounds and heritage languages produced overall lower vowel durational variability than previously reported values of European American speakers. Contrastingly, Newman and Wu (2011) found that average prosodic rhythm in eight AAC speakers did not appear markedly different from those of other racioethnic backgrounds (White, Black, Hispanic). Only the two Chinese American men in the sample (as opposed to the two Chinese American women and all four Korean Americans) produced speech that was particularly low in durational variability. A separate small-scale study of four Chinese Americans report that some speakers demonstrated lower rhythmic variability and/or contextual flexibility based on interlocutor, suggesting that rhythm may be an ethnolinguistic feature available for indexing Chinese-American identity (Zipp & Staicov, 2016). Given these limited results, it is unclear whether reduced rhythmic variability is broadly leveraged as an AAC identity marker, but there may be support for it being part of a Chinese American ethnolinguistic repertoire.

3.2.3 The Current Study

Descriptions of AAC speech production contribute to the overall understanding of AAC speech as it may be cognitively represented. Since relatively little is known on the whole, the current study takes an exploratory approach, allowing for both quantitative and qualitative insights into sociophonetic variation of AAC English, including those produced by different AAC ethnic groups, that can inform future research. I specifically explore self-identified AAC speech production with an eye towards what it might tell us about perception of AAC speech.

Examining publicly-available naturalistic recordings as a source of AAC-identified speech could connect the results to more realistic settings. YouTube, an online video sharing platform, is a vast yet relatively untapped source of linguistic data (E. W. Schneider, 2016) which provides the ability to collect larger amounts of "naturalistic" speech data representing various contexts (for advantages of self-recordings, see Hall-Lew & Boyd, 2020) and much of this speech is (at least automatically) captioned. There are also disadvantages, including missing or difficult-to-obtain demographic speaker information and lack of control over audio quality. However, an increasing number of studies demonstrate that it is possible to use YouTube speech data to study phonetic variation and change, including context-based style-shifting (S. Lee, 2017), region-based articulation rate (Coats, 2020), and longitudinal second dialect acquisition (A. Cheng, 2022). This study serves as another demonstration of how YouTube may be used as a data source in sociophonetic research, especially given an exploratory approach.

In the current study, I present an exploratory corpus analysis of linguistic features in AAC speech production to add to the limited literature on this topic. To assess whether previous findings can be generalized to a different sample of AAC speakers, I examine the same phonetic features as Bauman (2016)—/oʊ/-backing, /oʊ/-monophthongization, and prosodic rhythm—in a corpus of YouTube videos, focusing on speech from Californian Asian Americans. While I take into account (pan-)ethnic identity as a potential factor in variability, many factors beyond categorical ethnic origins shape individual speech patterns, including the degree of ethnic identification and intersecting social identities. Thus, in addition to considering speech patterns across all AACs and different AAC ethnic groups, I consider the extent to which individual speakers utilize these phonetic features alone and in combination.

3.3 Methods

3.3.1 Data Source and Speakers

The speech data come from publicly-available self-recorded videos in video blog, or “vlog”, format posted to YouTube. The procedure for selecting speakers and speech samples occurred in several stages from January 2021 to June 2021. In the initial stage, a broad pool of self-identified Asian American speakers were identified for consideration via Asian American topic videos (e.g., “Growing up Asian

American" tag). A review of the video content for speaker identity (supplemented by other sources, such as YouTube channel "About" pages) revealed that a large number of speakers were from California and the majority presented as women. As a sample balanced for demographic background did not appear to be possible, I chose to limit my intended speaker sample to those who grew up in California (as self-identified by speakers), and visually presented as women (as judged by the author) in an attempt to reduce some degree of sociolinguistic variability for data interpretability.

In the next stage, the main search process, I expanded my search for self-identified Asian Californians to other types of videos with a similar format where they may share personal demographic information with their audience (e.g., "Get to Know Me" tag). Potential speakers were screened for demographic information as before, mainly via review of video content. At the same time, speakers were screened for inclusion based on whether speech from their videos were appropriate for phonetic analysis and whether there was enough speech to analyze. For each speaker, Growing up Asian American or Get to Know Me videos were screened first, and when deemed necessary, other vlog or Q&A format videos from the speaker's channel were screened, with preference given to videos of similar content. In general, videos were included only if the majority of the speech content involved one speaker addressing the audience in English, little to no background music or environmental noise, and acceptable audio quality. In addition, I aimed to collect at least 15 minutes of usable speech, estimated from video length (where videos involved mainly speaking) or timestamps of speaking portions (where videos had significant non-speech portions). If the speech duration threshold could not be reached, that speaker was excluded from the final sample. Finally, I conducted the same search for comparable videos from Californian speakers of other racioethnic backgrounds to serve as comparisons.

This proved to be difficult, as self-identification of race and ethnicity were relatively rare occurrences; the need to screen for regional origin further narrowed the pool significantly.

In the end, 28 Californian speakers (23 Asian American-identified, 5 non-Asian American-identified) were selected for analysis based on ethnicity and regional origin. In this sample, all Asian American speakers could be considered either East Asian or Southeast Asian, including the following self-identified ethnic identities: 7 Korean, 5 Chinese, 3 Vietnamese, 3 Taiwanese, 1 Chinese-Korean, 1 Filipino, 1 Malaysian, 1 Indonesian, and 1 Cambodian. The non-Asian American comparison speakers included two European/White Americans ('Danish-German', 'half Irish'), two Latin/Hispanic Americans ("Mexican", "Mexican-Panamanian"), and one with mixed ancestry ("half Hispanic" and "half White").

3.3.2 Data Processing

Following manual speaker and video selection, video captions and audio were downloaded (web scraped) and processed via LingTube, a Python pipeline developed to scrape and pre-process linguistic data from YouTube videos (L. S. P. Cheng & Kramer, 2021; Kramer, 2021). Captions were semi-manually corrected and time-aligned, then speech segments were identified via automatic forced alignment using the Montreal Forced Aligner (McAuliffe et al. 2017). At this stage, target vowel boundaries were hand-corrected and coded for errors or issues that may interfere with analysis (e.g., overlapping noises, creakiness).

For the purposes of the durational measures, target vowels included 250 consecutive vowels beginning from the first available video per speaker. In addition, to assess formant-based measures, target vowels included all primary-stressed tokens of /oʊ/, plus a subset of other primary-stressed reference vowels (/æ/, /i/, /u/, /ɑ/), aiming for at least 15 or more. In general, vowel boundaries were placed based on the start and end of (complex) periodicity and/or formants (particularly F2), as well as changes in waveform amplitude; in cases without a clear boundary, such as surrounding approximants or vowels, auditory impressionistic judgments were used to determine the boundaries. Vowels were only included when fully voiced, meaning syllabic consonants or devoiced vowels were excluded. In addition, filler planning words (e.g., *um*, *uh*, *so*) were excluded at this stage.

Following boundary correction, the duration and first three formants were extracted for each vowel via FastTrack (Barreda, 2021). Prior to formant processing, I conducted a manual check of a subset of each speaker's /i/ and /u/ tokens (roughly 10-15 each) to identify optimal maximum and minimum formant ceiling values for that individual. Then, FastTrack was run in Praat individually per speaker to collect formant estimates from the first three formants across the duration of the vowel. Rather than single point estimates, FastTrack provides the average (median) value within a certain number of user-specified time bins, resulting in more reliable estimates; in this case, I chose nine bins, such that the third, fifth, and seventh bins represent F1, F2, and F3 estimates at approximately 30%, 50%, and 70% into the vowel. Following extraction, formants were scaled using single-parameter log-mean normalization, selected because it is argued to better preserve meaningful phonetic variation (Barreda & Nearey, 2018; Barreda, 2021).

To process vowels for formant analysis, the target vowel data were filtered to remove tokens with errors, voiceless production, or other issues that would interfere with formant measures (e.g., creaky, noisy) as coded during hand-correction. Tokens with very short durations (<40 ms for monophthongs, <75 ms for diphthongs) were also removed. Formant tracking outliers were estimated using the Mahalanobis distance (i.e., distance between points in a multivariate space) on F1, F2, and F3 values between 30% to 70% into the vowel. Those that exceeded three times the standard deviation from the mean were removed. Due to the limited number of words with primary-stressed /oʊ/ produced by certain speakers, selected function words that contained the target /oʊ/ vowel and were judged by the author as non-reduced during hand-correction were retained for analysis (e.g., *most*, *over*, *no*); other function words were excluded. In addition, to reduce the coarticulatory effects of surrounding segments on the target /oʊ/ vowel, in particular an effect of fronting inhibition (e.g., Fridland & Bartlett, 2006; Labov et al., 2006), tokens where /oʊ/ was preceded by a glide or followed by a glide, nasal, or liquid were removed (e.g., *know yourself*, *older*, *home*). Finally, to reduce lexical/contextual influence, at most five tokens of any word (selected as the first five) were included in the final analyses. In the end, the dataset included a total of 1184 tokens of /oʊ/, where each speaker was represented by 24-76 tokens (M=42.29).

Two measures were taken from the formant dataset. First, to represent degree of /oʊ/-backing, the log-mean normalized nucleus F2 value was used. In this case, lower F2 values represent more backed /oʊ/, while higher F2 values represent more fronted /oʊ/. Second, to assess degree of /oʊ/-monophthongization, I use the normalized Euclidean distance (nED) between formants earlier and later in the vowel as a measure of vowel dynamism. Specifically, this measure takes the Euclidean distance of F1 and F2 between the timepoints at approximately 30% and 70% into the vowel, divided by the token's duration; this second step serves to normalize for durational effects of vowel trajectory, as longer vowels tend to show more movement overall. Larger values represent larger movement, or more diphthongization, while lower values represent more monophthongal vowel quality.

To process vowels for the durational analysis, a separate subset of data was constructed. Data from the first 250 consecutive vowels were filtered to remove vowels with forced-alignment errors or voiceless production, as coded during hand-correction. The total dataset included 6151 tokens, where each speaker was represented by 198-236 tokens (M=220). From these data, one aspect of speech rhythm is assessed, specifically vowel durational variability measured using the normalized Pairwise Variability Index for vowels (nPVI-V; Grabe & Low, 2002). I follow a modified procedure for naturalistic speech based on previous sociolinguistic work (Thomas & Carter, 2006; Bauman, 2016), but in this case included only voiced vowels as vocalic intervals and did not control for phrase final lengthening (following Grabe & Low, 2002; White & Mattys, 2007). The nPVI-V is calculated as the median of the differences between successive intervals divided by the sum of the same intervals (multiplied by 100). In this way, the nPVI-V assesses variability in duration across pairs of subsequent vowels. The measure results in one score per speaker between 0 and 100: Scores closer to 0 have less variability across vowel durations, while those closer to 100 have higher variability.

3.3.3 Data Analysis

To explore patterns based on the combination of all three features in a bottom-up manner, I conduct a hierarchical clustering analysis on the acoustic measures. Analysis was conducted in R following recommended clustering protocols (Hair et al., 2019).

The continuous variables for clustering were by-participant (n = 28) means for the three phonetic features of interest: /oʊ/ nucleus F2 values, /oʊ/ F1-F2 Euclidean distance values, and vocalic nPVI. Prior to clustering, variables were assessed as independent (i.e., no variables were highly correlated, set to a threshold of $r \geq 0.9$) and clustering tendency in the data was confirmed (Hopkins statistic=0.54, which is over the standard threshold of 0.5). In addition, Malahanobis distances were calculated to assess outliers prior to clustering; none were found, so all data was included in analysis. Raw scores were z-scored (centered and scaled), then converted to Euclidean distances.

The three scores per speaker (mean /oʊ/ nucleus F2, mean /oʊ/ nED, nPVI-V) were submitted to hierarchical agglomerative clustering via average-linkage method. This clustering approach starts with each case as individual clusters that are linked with the next most similar cluster in a stepwise fashion until one single cluster of all cases is reached. Thus, the number of clusters does not need to be predetermined. The average linkage method was selected as it finds clusters of any size or shape ("classes") and is less affected by outliers.

3.3.4 Predictions

Broadly, previous research suggests that AACs, as compared to non-AACs, may produce (a) more retracted /oʊ/, (b) more monophthongal /oʊ/, and (c) lower vocalic durational variability as markers of an Asian American identity (e.g., Bauman, 2016). However, other research suggests that there is likely ethnicity-based variation within the broader AAC category, although predictions for different ethnic groups are limited. Minimally, /oʊ/-backing and -monophthongization (lower F2 and nED) may be expected for Korean Americans, following the broader AAC prediction, while Chinese Californians may be expected to produce fronted /oʊ/ (higher F2), similar to reports for European Californians. In addition, lower vocalic variability (lower nPVI-V) may be expected for Chinese Americans in particular, while this may not be the case for Korean Americans. For reference, previous reports have suggested nPVI-V scores of roughly 50-55 for White and Black speakers and 30-43 for Hispanic speakers (Thomas & Carter, 2006), while AAC speakers have been reported to fall in the range of 33-53 (with an average of 43; Bauman, 2016). Although these predictions are provided for context, this corpus-based sample is not well-suited for confirmatory investigation as it is neither large nor balanced across groups; thus, my main goal is exploratory description across multiple levels, particularly at the level of the individual where I expect a high degree of variability.

3.4 Results

For each phonetic feature, I consider comparisons between (1) AAC and non-AAC speakers, (2) speakers of different ethnic groups, and (3) individual speakers. For ease of data description and visualization, in addition to the groups of "Korean", "Chinese", and "Vietnamese", the remaining speakers were placed into "East Asian" (Taiwanese, Chinese-Korean) and "Southeast Asian" (Filipino, Malaysian, Indonesian, Cambodian) categories; however, they are still identifiable by ethnicity via the speaker codes.

3.4.1 /oʊ/-Backing

The total number of /oʊ/ tokens was 1184 and the mean number of /oʊ/ tokens per speaker was 42 (range: 24-76). Figure 3.1 presents the mean normalized nucleus F1 and F2 values of /oʊ/ per speaker, where color indicates AAC (red) and non-AAC (blue) speaker categorization. Figure 3.2 displays the same F2 values colored by ethnicity groupings. In line with typical formant plots, the x-axis is reversed such that lower F2 values, or more backed tokens, are located on the right. As the feature of interest is /oʊ/-backing, I focus on F2 measurements.

At the highest level of categorization, non-AAC speakers on average produced /oʊ/ with higher F2 values (M = -0.03, SD = 0.17), representing a fronter articulation, compared to AAC speakers who as a group produced lower average F2 values (M = -0.07, SD = 0.18), representing more retracted /oʊ/. Although this direction of difference is consistent with the prediction that /oʊ/-backing may be an indicator of AAC identity, Figure 3.1 indicates a more complex picture. While the non-AAC benchmark speakers are highly consistent in terms of mean F2 values—

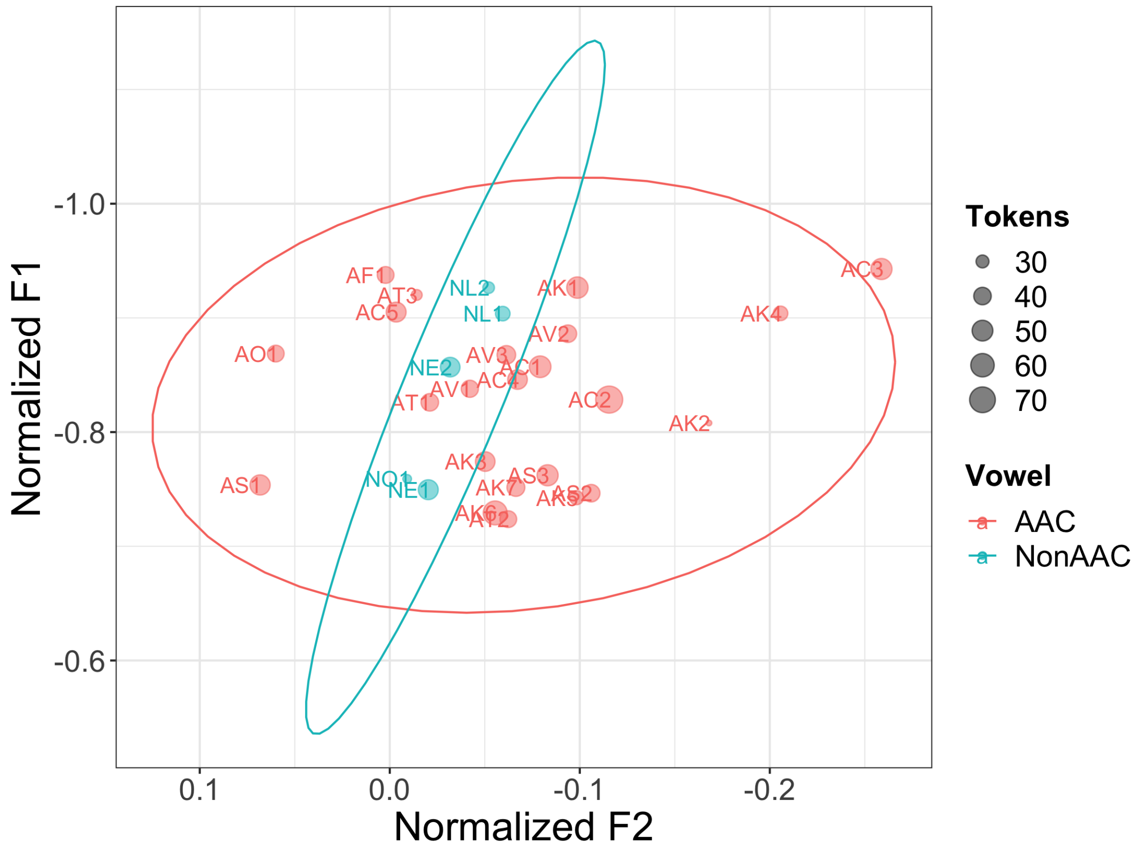

Figure 3.1: Log-mean normalized F1 and F2 values at /oʊ/ nucleus per speaker. The x-axis is reversed such that lower F2 values, representing more backed vowels, are located to the right.

even surprisingly so—AAC speaker means are highly variable. A majority cluster near the non-AAC speakers with slightly lower F2—as the category means would suggest—but there are also some speakers with very low F2 values (extremely retracted) and another set with very high F2 values (extremely fronted), even beyond those of the non-AAC.

Taking a closer look at the data, some trending patterns emerge based on ethnicity. Table 3.1 presents mean F2 values by ethnicity grouping, where rows are ordered from highest to lowest values. Korean and Chinese ethnic groups produced notably lower mean F2 values, indicating relatively backed /oʊ/. By-individual examination of the data shows that this generally holds true across speakers, and

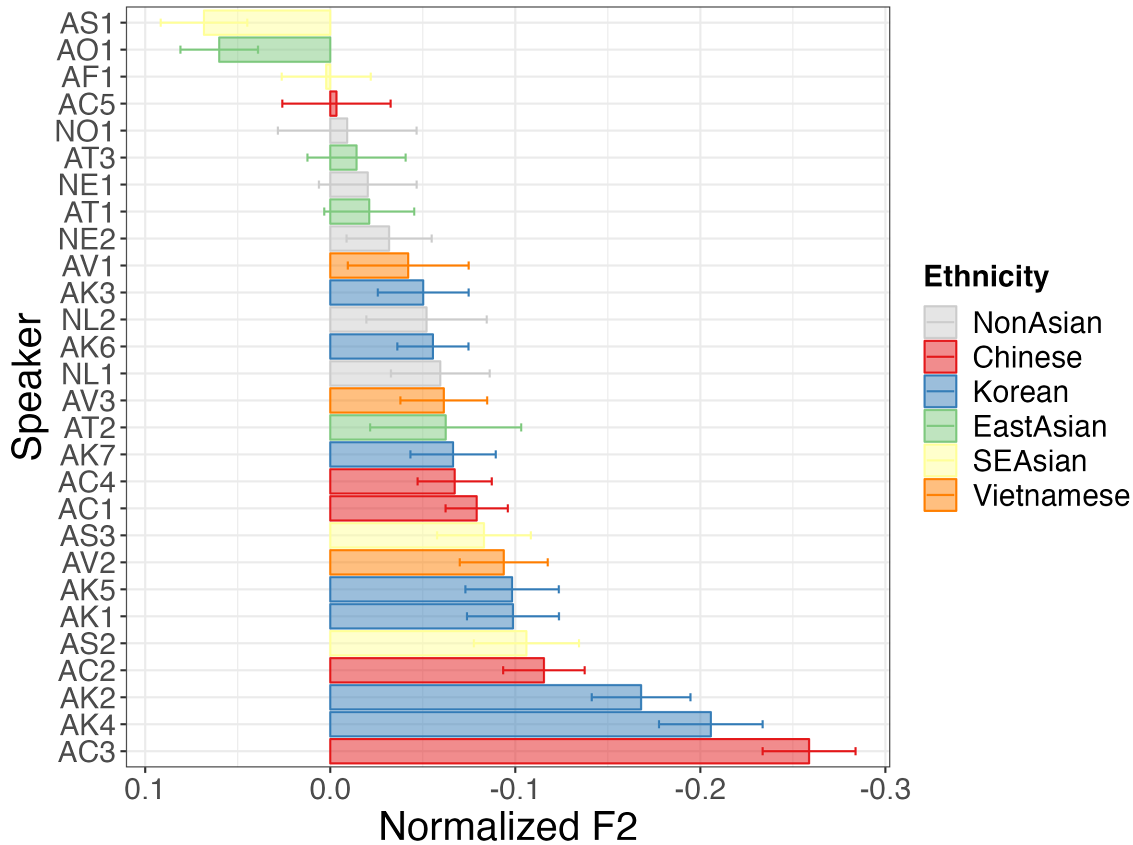

Figure 3.2: Log-mean normalized F2 values at /oʊ/ nucleus per speaker, presented by ethnicity. The x-axis is reversed such that lower F2 values, representing more backed vowels, are located to the right.

Ethnicity Group	Mean F2 (log-mean normalized)	SD
EastAsian	-0.009	0.181
SEAsian	-0.029	0.182
NonAsian	-0.034	0.168
Vietnamese	-0.066	0.170
Korean	-0.095	0.164
Chinese	-0.108	0.188

Table 3.1: Log-mean normalized F2 values at /oʊ/ nucleus, averaged by ethnicity grouping. The x-axis is reversed such that lower F2 values, representing more backed vowels, are located to the right.

that the three speakers with the most extreme F2 values in the low range are Korean (AK2, AK4) or Chinese (AC3).

The Vietnamese group appears to trend in the same direction of /oʊ/-backing, but examination of the individual means show that two of the three Vietnamese speakers have very similar F2 values to non-Asian speakers; more data from this ethnic group will be needed for making any conclusions. On the other hand, the Southeast Asian and East Asian group means suggest higher F2 values than non-Asian speakers—extremely so for the East Asian group—opposite to predictions. This pattern appears to be driven by two speakers (AS1, AO1) with exceptionally high relative F2 while the other speakers are either very similar to the benchmark speakers or actually rather backed (AT2, AS3, AS2). Because these analysis categories grouped together individuals with various ethnic backgrounds, it is difficult to make any interpretations as to why. Suffice to say, however, that this demonstrates the large range of variability across AAC speakers with regards to /oʊ/-backing.

In sum, these findings are somewhat mixed relative to past literature. In this particular sample of Californian speakers, Asian American speakers as a group produced /oʊ/ that was typically further back than non-Asian American speakers, driven in particular by many of the Korean and Chinese Americans. These findings are consistent with findings of /oʊ/-backing as an Asian American identity marker (Bauman, 2016) and a Korean ethnic marker (L. Jeon, 2017; A. Cheng, 2020) but somewhat inconsistent with prior reports of Chinese Californians tending to front /oʊ/ (Hall-Lew, 2009; A. Cheng et al., 2016). At the same time, not all Asian American speakers showed the overall backing tendency, underscoring high variability across individuals as reported previously and as expected due to the diversity within this sample.

3.4.2 /oʊ/-Monophthongization

Figure 3.3 presents the mean normalized F1 and F2 values at 30% (origin) and 70% (arrowhead) through the vowel, where color indicates AAC (red) and non-

AAC (blue) speaker categorization. Figure 3.4 displays the nED values representing the vowel movement over time, colored by ethnicity groupings.

As a group, non-AAC speakers produced /oʊ/ with lower nED values (*M* = 1.41, *SD* = 0.77), representing less movement or relatively monophthongal articulation, compared to AAC speakers who as a group produced higher average nED values (*M* = 1.67, *SD* = 0.99), representing more movement or relatively diphthongal articulation. This direction of difference is in fact opposite to the prediction that /oʊ/-monophthongization may be an indicator of AAC identity.

Examining these data by ethnicity provides some more nuance to these

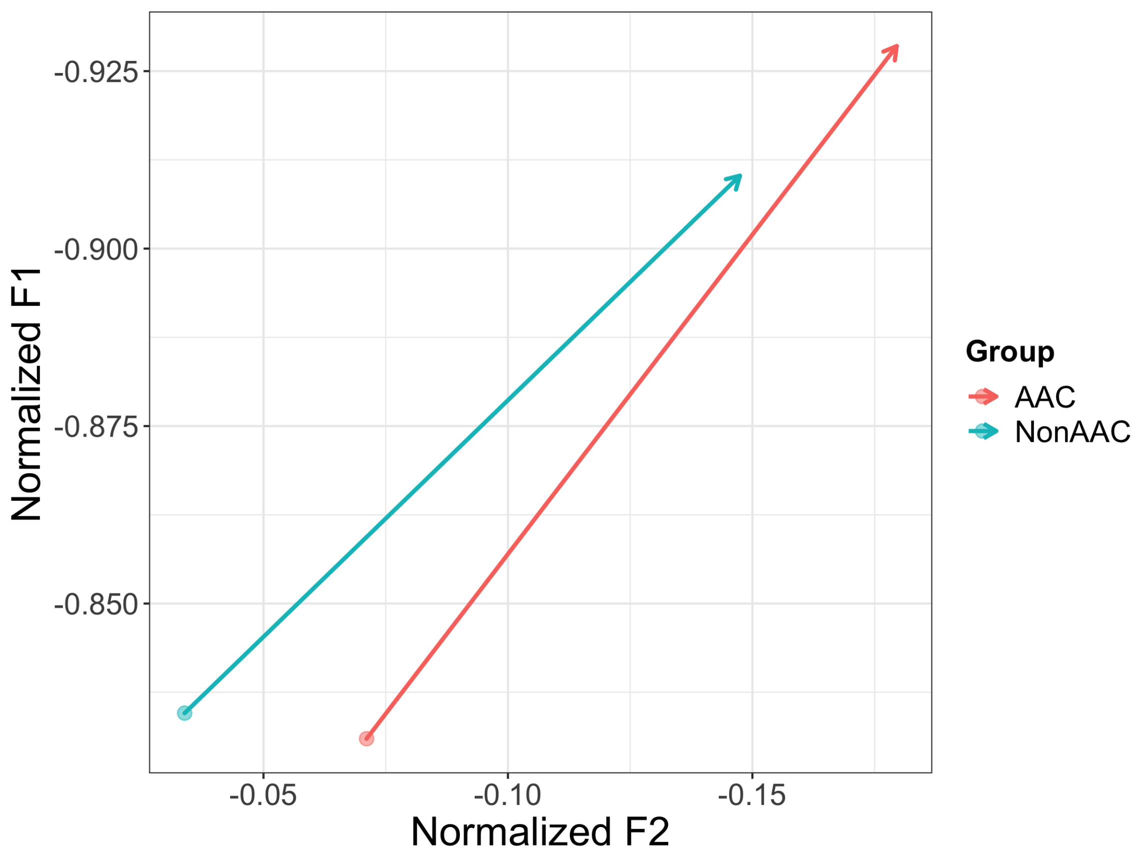

Figure 3.3: Log-mean normalized F1 and F2 trajectories for /oʊ/, plotted by AAC and non-AAC group. The origin point represents formant values at 30% into the vowel and the arrow point represents values at 70% into the vowel. The x-axis is reversed such that lower F2 values, representing more backed vowels, are located to the right.

findings. Table 3.2 displays mean nED values by ethnicity grouping, where rows are ordered from highest to lowest values. Non-Asians as a group indeed have the lowest group average for diphthong movement, making them the most monophthongal, contrary to predictions. They are followed by the Korean American group, of whom three (AK1, AK7, AK2) are among the most monophthongal individuals, including the speaker in this sample with the lowest mean nED (*M* = 1.18, *SD* = 0.57). However, this feature was variable with this ethnic group as several other Korean speakers are more moderate, while another (AK5) was actually the most diphthongal of all individuals (*M* = 2.22, *SD* = 1.29).

Southeast Asian speakers on average had the highest nED values, which is gen-

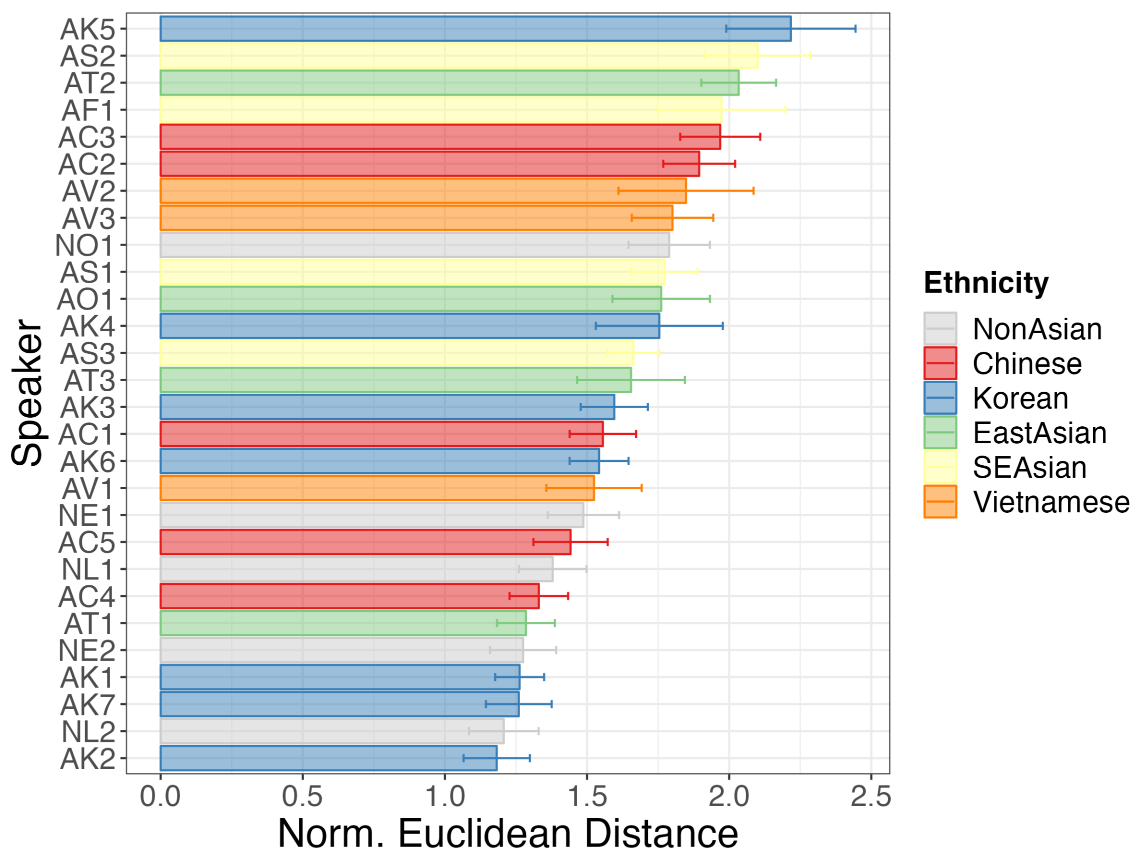

Figure 3.4: Normalized Euclidean Distance (nED) based on log-mean normalized F1 and F2 values for /oʊ/ from 30% to 70% through the vowel. Color represents ethnicity.

Ethnicity Group	Mean Euclidean Distance (normalized)	SD
SEAsian	1.858	1.017
Vietnamese	1.728	1.183
EastAsian	1.680	0.912
Chinese	1.670	0.968
Korean	1.530	0.929
NonAsian	1.411	0.774

Table 3.2: Normalized Euclidean distance between 30% and 70% across /oʊ/, averaged by ethnicity grouping.

erally observed in the individual data as well. Finally, Chinese, East Asian, and Vietnamese group means were somewhere in the middle in terms of /oʊ/ formant movement. Inspection of the individual scores suggests that production of this feature was highly variable within these groups such that no particular pattern stands out as characterizing speakers of these ethnicities. For example, while two Chinese American speakers (AC4, AC5) have relatively low nED, or more monophthongal pronunciation, two other speakers (AC2, AC3) have relatively high nED, producing quite diphthongal /oʊ/. Overall, individual variability seemed to be more prominent than any patterning by specific ethnic identity.

These findings appear rather inconsistent with prior research. In the current sample of Californian speakers, Asian American speakers as a group produced /oʊ/ that was typically more *diphthongal* than non-Asian American speakers. These findings directly contrast with reports of /oʊ/-monophthongization being used as a marker of Asian American identity (Bauman, 2016) with links to Japanese American ethnicity specifically (D'Onofrio & van Hofwegen, 2020). However, Korean Americans seemed to be more consistently monophthongal than other Asian American speakers, which falls in line with non-significant trends reported in previous research (D.-E. Kim, 2021). Admittedly, previous studies of monophthongization in AACs have been sparse, and considering the high variability in the current data, an interpretation of a link between formant trajectory in /oʊ/ and AAC ethnic identity is at best inconclusive.

3.4.3 Speech Rhythm

A total of 6151 vowel tokens, around 200 per speaker (*M* = 220, range: 198-236), were used to calculate nPVI scores. Figure 3.5 displays the vocalic nPVI values per individual speaker, presented by ethnicity. The full set of scores ranged between 36 to 60 (*M* = 48.75, *SD* = 5.37). Mean nPVI score across the 23 AAC speakers was 49.34 (*SD* = 5.36), spanning the full range of variability, while the 5 non-AAC speakers' nPVI mean was 46.02 (*SD* = 5.07), spanning a relatively wide range between 40 to 53.

While the average nPVI is comparable, or even lower for non-AACs (contrary to predictions), the breakdown by ethnicity and individuals shows some structured variability. Table 3.3 presents mean scores by ethnic grouping. As a group, Chinese American speakers produced speech with the lowest nPVI (*M* = 44.73, *SD* = 4.90), including the speaker (AC4) with the lowest overall nPVI (36.20). They are followed next by Vietnamese Americans (*M* = 45.57, *SD* = 2.52). Non-Asians (*M* = 46.02, *SD* = 5.07) are in the middle, though the variability is quite high, with speakers spanning the low end of the scale (NL1, NE1) and the high end (NE2). On the other hand, results show that Korean American speakers tend to have consistently higher nPVI (*M* = 50.26, *SD* = 3.21), as do the speakers categorized as Southeast Asians (*M* = 51.47, *SD* = 6.22) and East Asians (*M* = 54.17, *SD* = 5.11).

Overall, the current vocalic nPVI values for AAC speakers fall in a somewhat similar range to those reported in Bauman (2016) (33-53, *M* = 43), though both the lower and upper bounds are shifted upwards. This also means that AACs cover a range corresponding to benchmarks for Hispanic (30-43) and European (50-55) Americans (Thomas & Carter, 2006), as well as including Newman and Wu's (2011) low-end range for Chinese American men (35-45). Moreover, the data

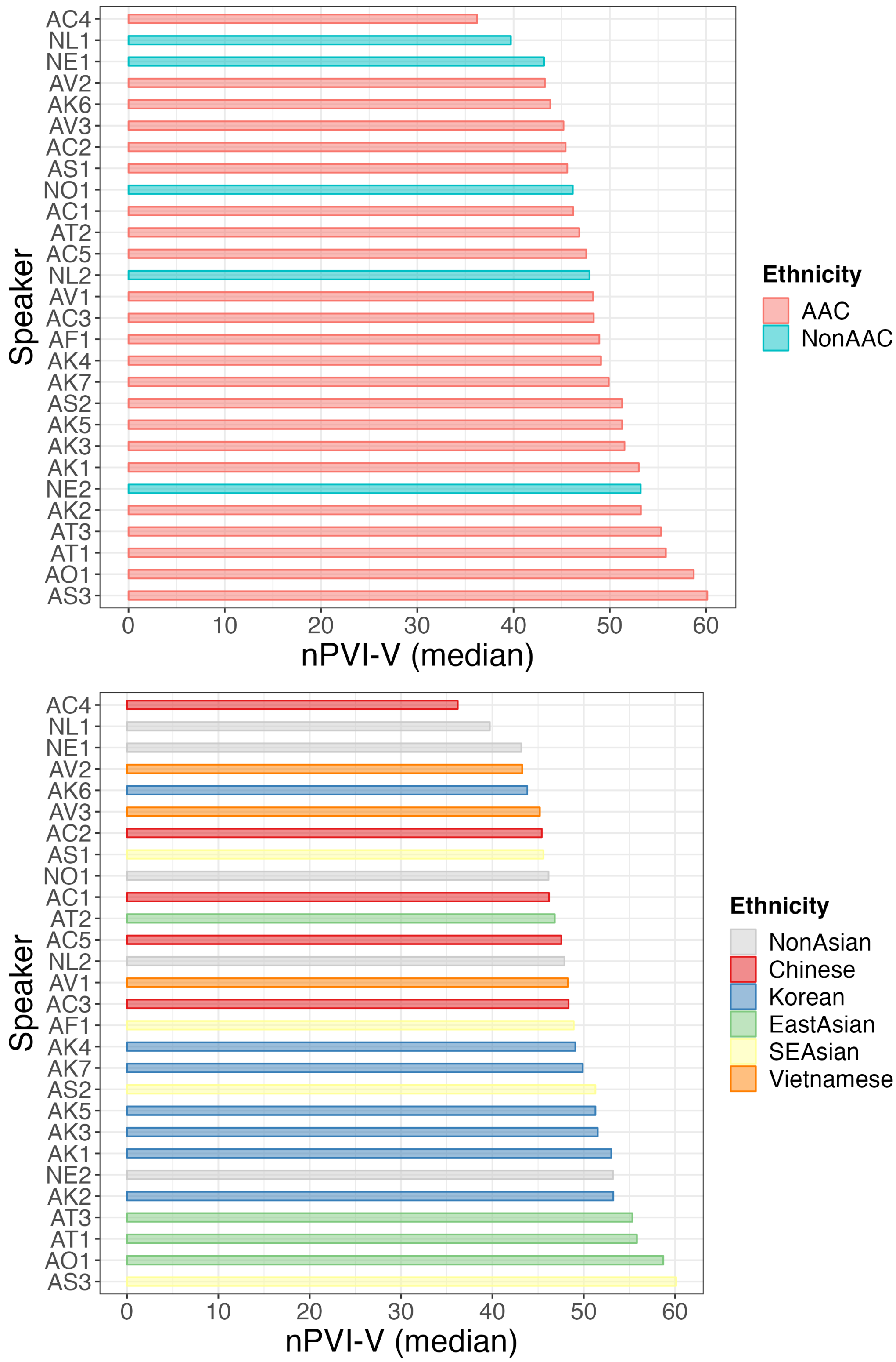

Figure 3.5: Median vocalic nPVI scores per speaker, where color represents AAC or non-AAC group (top) or specific ethnicity (bottom).

Ethnicity Group	Mean nPVI	SD
Chinese	44.73	4.90
Vietnamese	45.57	2.52
NonAsian	46.02	5.07
Korean	50.26	3.22
SEAsian	51.47	6.22
EastAsian	54.17	5.12

Table 3.3: Normalized Pairwise Variability Index (nPVI) for vowel durations calculated as the median per speaker and averaged by ethnicity grouping.

are descriptively similar to Bauman's in that AACs define the lowest and highest values while non-AAC speakers at the same time present relatively high variability as well. However, given the small non-AAC sample size, the variability is too high to be interpretable as a useful reference point. That is, although in this sample, Latina Americans had relatively low average nPVI (M = 43.80) and European Americans had higher nPVI (M = 48.18), scores are not consistent within group as both included one speaker with very low nPVI compared to the others with moderate or high nPVI. Instead, I focus on the variation within the set of AAC speakers.

Interestingly, we find that the descriptive ethnic group patterns are relatively consistent with the limited literature on this topic. Specifically, the potential difference between some Chinese and Korean Americans (e.g., Newman & Wu, 2011) in prosodic rhythm was borne out in the current data: Chinese American speakers tend to produce speech in the lower range of this sample (36.20-48.33) while Korean American speakers tend to produce speech with higher vowel variability—aside from AK6 (43.81), the remaining six speakers produced nPVI values between 49.08-53.20. In addition, the current data suggest that Vietnamese American speakers may generally be similar to Chinese Americans in prosodic rhythm, while the Taiwanese American speakers (grouped under East Asian) appear similar to Korean Americans. Future research would be needed to confirm these trends.

In all, given the wide range of nPVI score within the AAC sample, it seems that vocalic durational variability is not necessarily a general feature used by speakers as an index of AAC identity but instead appears tied to more specific ethnolinguistic repertoires. If so, on the part of listeners, prosodic rhythm could be a feature that cues them to interpret a speaker as having a particular ethnic identity (say, Chinese). Still, it is possible that prosodic rhythm could be part of a pool of features associated with AAC speakers in general; in other words, if listeners extrapolate from specific ethnic groups or speakers to form a representation for AAC (or "Asian") speech (e.g., "Mock Asian"; Chun, 2004), rhythm could be associated with a broader range of AAC identities, even those whose speech production in reality do not generally align.

3.4.4 Combination of Features

In order to select the most appropriate clustering solutions to discuss, the *NbClust* (Charrad et al., 2014) package was used to assess 30 indices of cluster validity which assess features such as cluster cohesion and separation. The outcomes indicate both 2 and 4 clusters (7/30 indices) as preferred options, followed by 8 clusters (5/30 indices). In order to provide a broader exploratory picture of the feature combination patterns in these data, I discuss the results of these top three cluster solutions. For each solution, I profiled the clusters by inspecting distinctiveness and means. Tables 3.4, 3.5, and 3.6, display the by-feature cluster means based on the original units of each measure. Figures 3.6, 3.7, and 3.8 visualizes the same patterns of cluster means by plotting z-scores (in order to compare across measures) by variable along with presenting individual z-score patterns within each cluster. These plots allow us to compare the relative differences between clusters and individuals across all three features at once.

In the 2-cluster solution (as well as the others), two speakers (AC3, AK4) are separated into their own cluster, Cluster A, characterized by particularly low /oʊ/ F2, slightly above average /oʊ/ nED, and fully average nPVI. In other words, one Chinese American and one Korean American speaker are clearly distinguished from the majority of speakers in this sample by extreme /oʊ/-backing (and possibly somewhat more diphthongal /oʊ/). This aligns with the independent examination of /oʊ/-backing results but provides the additional insight that those particular speakers are extreme mainly on this specific feature. The other cluster contains all remaining speakers and evidently involves extreme variability, which the following clustering solutions help shed light on.

The 4-cluster solution retains Cluster A and breaks up the other speakers into

Cluster	n	/oʊ/ nucleus F2	/oʊ/ nED	nPVI-V
A	2	-0.232	1.861	48.70
B	26	-0.051	1.609	48.75

Table 3.4: Cluster size and means per feature for the 2-cluster solution.

Cluster	n	/oʊ/ nucleus F2	/oʊ/ nED	nPVI-V
A	2	-0.232	1.861	48.70
B	3	-0.012	1.693	58.05
C	9	-0.053	1.937	47.10
D	14	-0.058	1.381	47.82

Table 3.5: Cluster size and means per feature for the 4-cluster solution.

Cluster	n	/oʊ/ nucleus F2	/oʊ/ nED	nPVI-V
A	2	-0.232	1.861	48.70
B	2	0.022	1.707	57.02
C	1	-0.083	1.663	60.13
D	6	-0.089	1.982	47.20
E	3	0.020	1.844	46.88
F	11	-0.047	1.403	49.12
G	1	-0.167	1.182	53.23
H	2	-0.063	1.354	37.96

Table 3.6: Cluster size and means per feature for the 8-cluster solution.

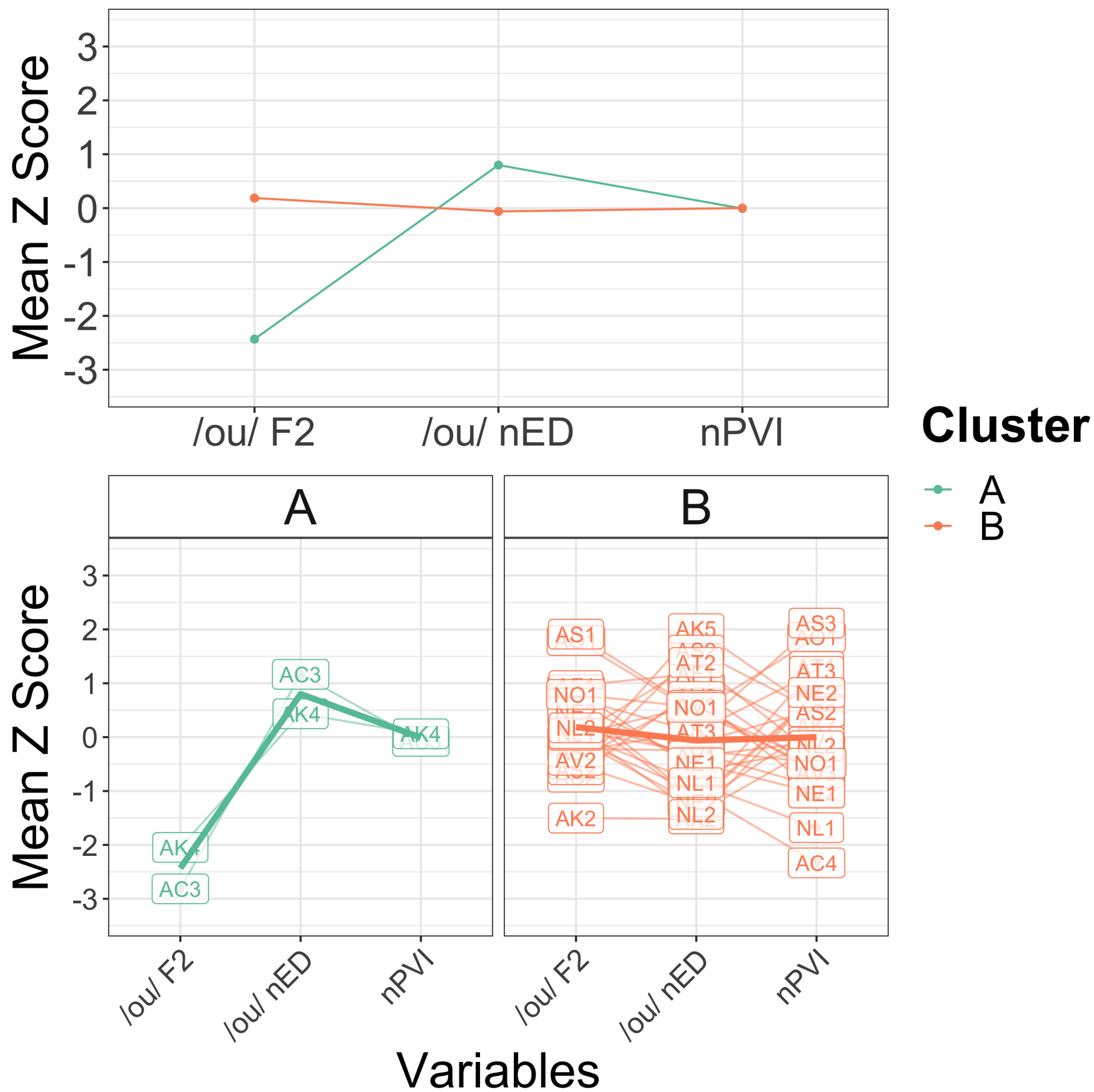

Figure 3.6: Top: Cluster z-score means per feature for the 2-cluster solution, colored by cluster. Bottom: The same data (bold line) are faceted by cluster and underlaid with individual z-score means (lines labeled with speaker) to represent cluster size and variability.

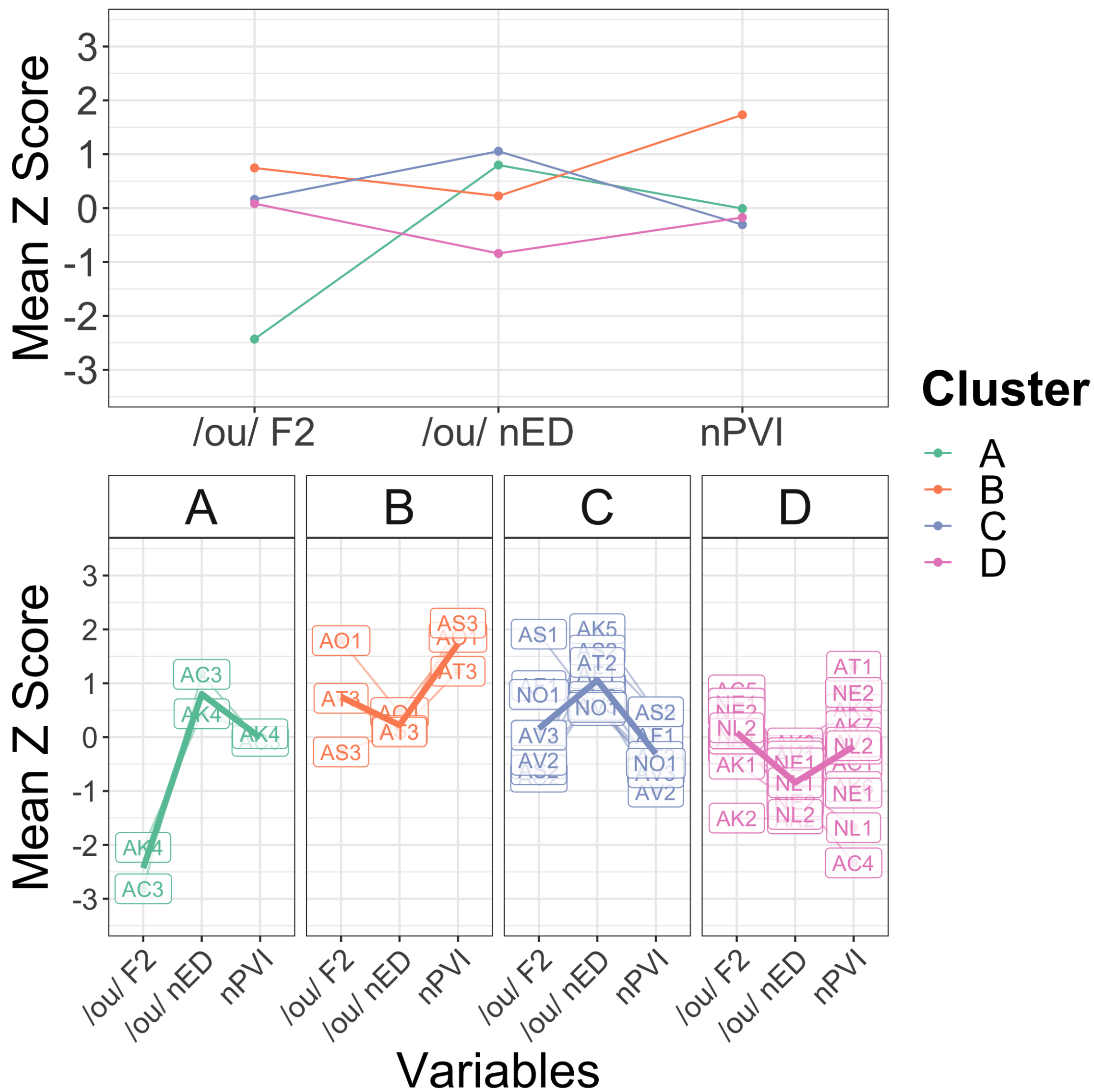

Figure 3.7: Top: Cluster z-score means per feature for the 4-cluster solution, colored by cluster. Bottom: The same data (bold line) are faceted by cluster and underlaid with individual z-score means (lines labeled with speaker) to represent cluster size and variability.

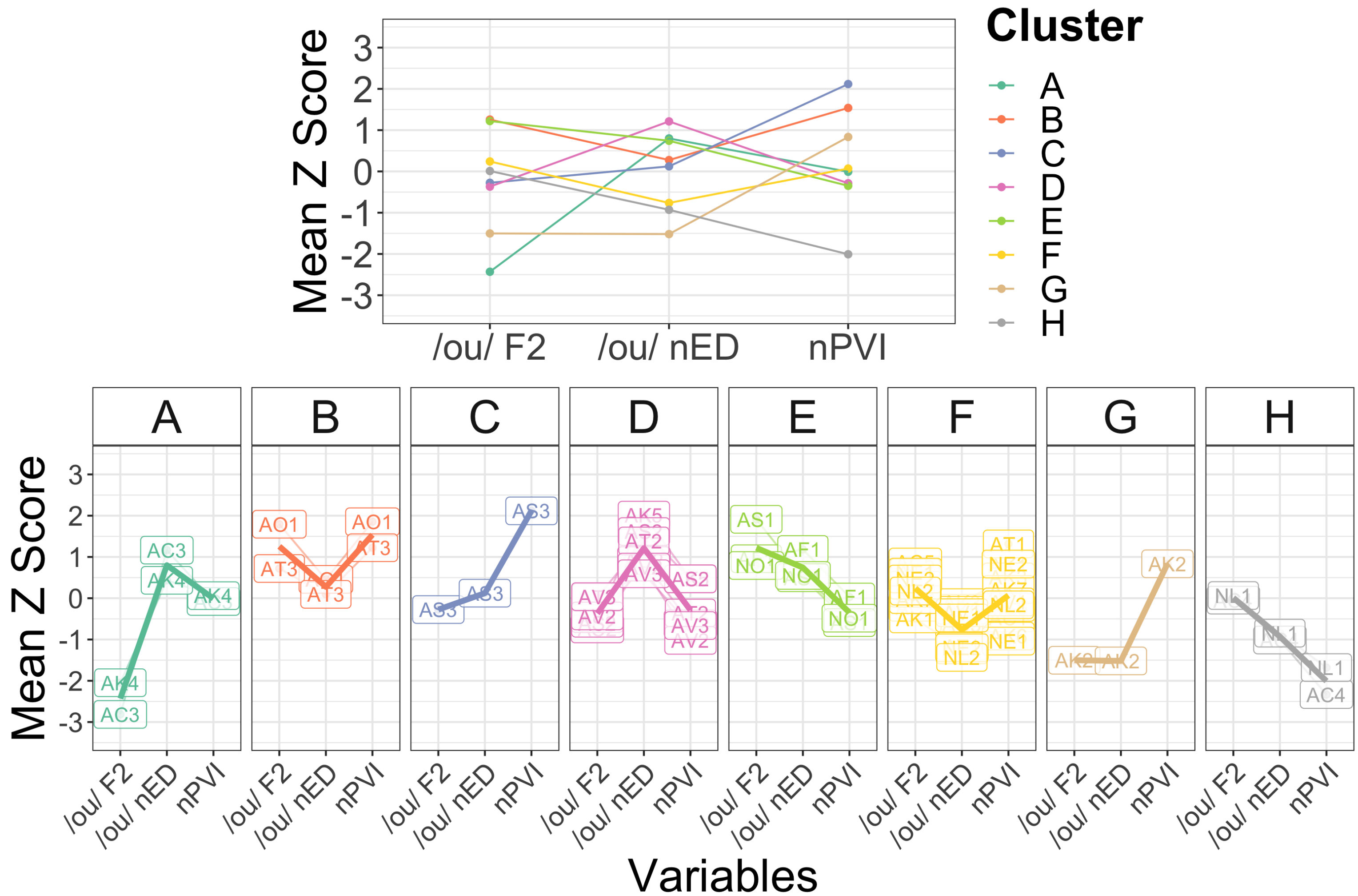

Figure 3.8: Top: Cluster z-score means per feature for the 8-cluster solution, colored by cluster. Bottom: The same data (bold line) are faceted by cluster and underlaid with individual z-score means (lines labeled with speaker) to represent cluster size and variability.

three smaller clusters. On average, each has one distinctive feature: Cluster B displays higher than average nPVI, Cluster C has relatively high nED scores, while Cluster D correspondingly has relatively low nED scores. However, each cluster still contains much variability; particularly clear is the variation of nPVI across speakers in the largest Cluster D. In terms of ethnic group composition, not many obvious consistencies arise. While it may be notable that the two Latina and two European Americans are contained within Cluster D (four of five non-AACs), as are a majority of Korean Americans (five of seven), this appears to straightforwardly reflect the lower average nED of these groups as reported in the independent analysis.

Lastly, the 8-cluster solution further divides the three larger clusters to reveal distinct feature combination patterns. Rather than one cluster of high nPVI, there are two: Cluster B (AT3, AO1) with both higher relative F2 (fronted /oʊ/) and nPVI (more rhythmic variability) versus Cluster C (AS3) with relatively average values other than high rhythmic variability. These represent some of the individuals with extreme values but no particular ethnicity-related pattern stands out. Similarly, instead of one high-nED cluster, there are two: Cluster D with only distinctly high nED and Cluster E which has relatively high nED but stands out with even higher average F2 for /oʊ/. So, while Cluster D contains six speakers with rather diphthongal /oʊ/ only, Cluster E's three speakers (AS1, AF1, NO1) have both highly fronted and diphthongal /oʊ/— a combination that is notably opposite to previous predictions about generally AAC-associated features, despite that two Southeast Asian Americans produce this pattern.

The last set of speakers all produced relatively low nED (i.e., monophthongal /oʊ/). However, the three sub-clusters reveal quite different feature combinations. Cluster F, composed of the majority of speakers in any single cluster (including 3 of 5 non-AACs and 4 of 7 Korean Americans), demonstrates mainly a relatively monophthongal production of /oʊ/. In contrast, Cluster G represents a single Korean American (AK2) who uniquely produces low F2 and nED for /oʊ/ and somewhat high nPVI. Interestingly, this pattern—backed and monophthongal /oʊ/ paired with high durational variability—mimics the Korean group results, meaning AK2 in a sense represents the "average Korean American speaker" in this sample, but is the only one who clearly shows this pattern (note, however, that other speakers may follow this pattern but produce less extreme values, thus ending up grouped with other less extreme speakers). Finally, Cluster H represents two speakers (AC4, NL1) who are very similar, producing low nED and extremely low nPVI. These are the two speakers—one Chinese American and one Mexican-Panamanian American—on the lowest end of the durational variability continuum, which aligns with predictions relating to ethnicity; these clustering results additionally show that they also produce more monophthongal /oʊ/ with roughly average backing (which means it is slightly retracted relative to European American speakers).

In sum, clusters do not reveal any obvious patterning by AAC or ethnicity categories in terms of combinations of /oʊ/-backing, monophthongization, and rhythmic variability. Rather, what stands out is individuals of various AAC ethnic backgrounds producing particularly high or low values on the various features, some of which align very well with AAC predictions (e.g., AK2 in Cluster G, AC4 in Cluster H) while others somewhat align (e.g., AC3 and AK4 in Cluster A) or do not align at all (e.g., Clusters B, C, D and E).

3.5 Discussion

This study examined three selected phonetic features in a corpus of Californian Asian American speech drawn from YouTube videos. The observations from this analysis both converge with and diverge from previous research on AAC

English speech. As predicted, Asian Americans as a group exhibited more retracted /oʊ/—led in this sample by Chinese and Korean Americans—compared to a smaller set of non-Asian Americans. There was also a group difference in /oʊ/ formant trajectories, but Asian Americans generally produced more *diphthongal* /oʊ/ than non-Asian Americans, contrary to expectations; although the subset of Korean Americans did present more monophthongal /oʊ/, that was in fact highly similar to the production of non-Asian Americans and unlike the majority of Asian Americans, raising questions about the characterization of /oʊ/-monophthongization as an AAC-associated feature. Prosodic rhythm involved a wide range of values within both Asian and non-Asian American groups, again suggesting a lack of general AAC association, but there was a consistency to the ethnolinguistic pattern of Chinese Americans exhibiting lower vowel durational variability while Korean Americans were the opposite, aligning with limited prior reports.

Taken together, there are more differences among those with different ethnic origins than similarities. For example, while at the group level Chinese and Korean Americans are similar on /oʊ/-backing, they differ in /oʊ/-monophthongization and prosodic rhythm. Moreover, clustering speakers based on these three features did not result in consistent groupings by ethnicity, but instead found individual AAC speakers producing many different combinations of features in ways that both matched and mismatched prior expectations of AAC speech. This also suggests that individuals did not generally produce patterns matching ethnic group averages. Overall, this highlights the high individual variability within both pan-ethnic and specific ethnic categories, which is a common observation in prior production (e.g., Newman & Wu, 2011; Bauman, 2016) and perceptual identification studies (e.g., Hanna, 1997; Newman & Wu, 2011; P. Wong & Babel, 2017; Nagy et al., 2020). This links to the idea of speakers having flexible use of linguistic resources within an ethnolinguistic repertoire (Benor, 2010) as well as the notion of listeners having access to a "constellation" of AAC-associated features—rather than any single feature—that inform perceptual identification (Newman & Wu, 2011); that is, speakers do not all use the exact same features but may draw from a similar feature pool.

One potential implication of such within-group variation is that different listeners could have rather different cognitive representations of AAC speech depending on their experience with not only different AAC ethnic groups but also individual AACs. For example, a person with greater exposure to Korean Americans or to certain AAC speakers who produce particularly backed /oʊ/ may have more robust /oʊ/-backing exemplar associations for "Asian" as a linked category. The weighting of this association could be further enhanced by awareness. As a result, this person should be more likely to leverage /oʊ/-backing as a cue to AAC speech compared to another person with limited exposure to Korean Americans or to extreme /oʊ/ backers in their social networks.

Nevertheless, this study examined only three possible phonetic features linked to AAC speech. Countless other features could serve as consistent perceptual cues to AAC identity, including those that might be found in production across multiple major ethnic groups or those that are tied to particular subcategories or types of AACs. In addition, features not found in production but which are present in ideologies of AAC speech may play an important role. More targeted research will be needed to determine what those features are, if they exist, and to what extent they are associated with AACs.

3.5.1 Limitations of the Dataset

There are a number of obvious limitations to this analysis, many stemming from the choice of data source and resulting lack of control. While YouTube videos are a plentiful source of a publicly available speech from a diverse population, audio data from YouTube—particularly when comparing across speakers—is necessarily noisy due to unknown and/or variable consequences of recording environment, equipment, audio compression, and so on (see, e.g., Sanker et al., 2021). This means that the outcomes of phonetic analysis may not be highly reliable, possibly leading to weaker or spurious effects; caution is warranted in interpreting these results pending replication within more controlled conditions. Nevertheless, the fact that descriptive results for /oʊ/-backing and prosodic rhythm were in directions consistent with prior evidence of Asian American speech strengthens the view that the data, while noisy, are at least somewhat representative.

Additionally, despite the amount and range of content posted to YouTube, selection of speakers and videos for phonetic analysis was in fact rather difficult. First, finding speakers with particular self-reported identities required expending vast amounts of time. In the first place, this information may not be shared by the YouTuber at all, and if it exists, the information must generally be gathered from video content, which is not easily searchable. Even if the information is found, it can be in varying levels of detail, which is out of the researcher's control. Moreover, this study had constraints of both ethnicity and region of origin, meaning that if only one was known, the speaker had to be excluded. On top of all this, the audio must be usable for phonetic analysis, and in fact a large proportion of videos contain background music or noisy environments, making it unfit for analysis. This also led to more varied speech contexts to be included across speakers rather than the same genre of video. Thus, this analysis effectively involved a convenience sample that was both small and highly diverse, making it difficult to form conclusions or generalize to the broader ethnic group or AAC population. For the same reasons, there was a very low number of comparison speakers included; these were even more challenging to identify as ethnicity is not often a topic of discussion.

More generally, use of YouTube data on its own means that demographic or personal background for each speaker must be coded from existing self-published materials. As this information is inconsistent across speakers and often in insufficient detail, other relevant social factors, such as age, ethnic orientation, and other languages known, cannot be included in the analysis for further consideration. In particular, ethnic orientation could well contribute to variability in the production results (e.g., L. Jeon, 2017; Nagy et al., 2020; D.-E. Kim, 2021) as could the extent and type of language contact. Again, however, that consistent descriptive patterns were discernible despite the lack of control in speaker selection helps to support the current interpretation of the data.

Future work considering the use of YouTube as a source of sociophonetic data may wish to carefully assess whether it is viable for their question or topic from a practical perspective. At the same time, alternative strategies could be undertaken to offset some of the potential difficulties with this data source. For example, to address the lack of detailed speaker identity information, one might choose to conduct targeted analyses of specific speakers for whom there is much publicly available information (e.g., A. Cheng, 2022) or contact the speakers directly for self-reported information.

Overall, given the particular sample analyzed in this dataset, we cannot conclusively contextualize the AAC results within the broader Californian population. At the same time, this study found that YouTube videos provided a rich data source for an exploratory approach to AAC speech production that allowed for description of sociophonetic variation within and across ethnic groups. Findings based on the relatively naturalistic nature of the data may also serve as a link to more realistic sociolinguistic contexts that complement other methodological approaches. In this way, YouTube-based data can be used as a jumping off point for future research that helps to build a fuller picture of particular ways of speaking and how they may be cognitively represented by others.

3.5.2 Conclusion

This study, together with previous literature, provides a grounding for how Asian Americans may be utilizing certain linguistic features in their speech. However, it does not alone allow us to directly connect these findings to listener expectations of AAC speech, particularly given the diversity of actual production patterns in such a sample. For example, although Chinese and Korean Americans may produce a relatively backed /oʊ/, is this a feature that listeners attend to in the speech they hear? If it is in cognitive representations, do listeners use it as a cue for Asian American speaker identity in perception? Other details of such a sociophonetic association could also be explored in more depth, including those probed by the following questions. Is /oʊ/-backing associated to a greater degree with certain types of Asian Americans, such as those who are thought to produce extremely backed /oʊ/? Whether listeners leverage this phonetic detail in perceptual behavior or not, is /oʊ/-backing a feature that is consciously accessible and available for discussion? If so, how strong or widespread is such awareness across individual listeners? To both further contextualize these production results and directly assess one form of listener expectations of AAC speech, I next gather perceptual judgments of speaker ethnic identity from Californian listeners, along with metalinguistic data on ideologies about AAC speech.

CHAPTER IV

Exploring Perception and Ideologies of AAC Speech

4.1 Introduction

Building off of Chapter III, this chapter presents a study examining one phonetic feature identified as present in AAC production—/oʊ/-backing—in perception and ideologies of AAC speech. I present a preliminary investigation of whether /oʊ/-backing may influence perceptual identification of AAC talkers, and ask to what extent this feature and other potential features are present in awareness within ideologies of AAC speech. To add context to these findings, I explore ideologies about social personae of AACs and how these may be linked to different linguistic expectations. Together, these results contribute to a more nuanced and ideologically-informed understanding of sociophonetic representations of AAC speech styles, which in turn informs in-the-moment linguistic expectations of AAC speech from the perspective of listeners. Broadly, it presents an approach to sociophonetic representations that begins to take into account which features might be more perceptually impactful due to conscious awareness, as well as which types of people might be more closely tied to certain features as a function of personae-based socioindexical associations.

4.2 Background

While sociolinguistic production research on AAC populations has been steadily growing especially over the last decade, only a handful of perceptual studies have examined AAC ethnic identification. These are broadly similar studies conducted in different locations across the U.S. and Canada where listeners are presented with short samples of speech and are asked to decide their impression of the speaker's racial/ethnic background, among other speaker characteristics. In terms of theoretical framing, these studies take two main approaches: (1) in the context of a general "Asian American" category, "Do some speakers 'sound Asian' to listeners and if so, what linguistic features give rise to it?" (Hanna, 1997; Newman & Wu, 2011; A. Cheng & Cho, 2021); and (2) in the context of specific ethnic groups, "Do listeners have knowledge of and use ethnolinguistic variation to make sense of speaker identity in perception?" (P. Wong & Babel, 2017; Nagy et al., 2020).

Across all studies, the ability to reliably—sometimes with high confidence and accuracy—identify AAC speakers of English certainly seems to exist, implying the existence of linguistic cues to AAC identity. At the same time, another consistent finding that tempers the previous is that both speaker and listener populations show large amounts of variability: not all speakers are identifiable as "Asian", and not all listeners are able to identify speakers as "Asian". The former highlights the role of individual speaker factors like ethnic orientation (Nagy et al., 2020), and the latter intersects with listener factors like experience or familiarity (P. Wong & Babel, 2017; A. Cheng & Cho, 2021). Further, clarity as to which linguistic or phonetic features contribute to signaling "Asian" or specific AAC ethnicity to listeners has yet to be achieved, despite various strategies to examine these questions in the context of perceptual results.

Thus, the question remains: What aspects of speech do listeners map to AAC identities, including broader and narrower ideological categories of AACs, and to what degree of awareness? In the following review, I first provide a brief summary of the perceptual literature's findings on this question. Then, I discuss some aspects of this question that have yet to be fully explored, but which are relevant for a full characterization of listener expectations of AAC speech, connecting back to the main questions of this dissertation about ideology, awareness, and expectations in sociophonetic processing.

4.2.1 Potential Cues to Asian Ethnic Identification

Previous research has explored potential perceptually-relevant features for ethnic identification via (a) metalinguistic commentary, (b) selective post-hoc phonetic analysis, and (c) fully systematic phonetic analysis. In the first case, studies that directly ask participants for metalinguistic discussion of AAC speech (using terminology such as "sounding Asian" or "Asian American accent") tend to highlight suprasegmental phonetic (e.g., prosodic, voice quality) or lexical features as cues that listeners consciously link to AAC speakers. For example, commentary on what makes somebody "sound Asian" includes mentions of high rising terminal ("upspeak"), "jerkier speech", and increased pauses or filler words (Hanna, 1997), as well as features such as higher pitch, "intonation", "softer tone", "vocal fry", and "clipped" words or syllables (A. Cheng and Cho, 2021). Segmental features appear less often in these data, but some have been discussed specifically in the context of Korean Americans (A. Cheng and Cho, 2021), including /ð/-stopping, /ɪ/ pronounced as [i], and /oʊ/ sounding "rounder" (potentially linked to backed and monophthongal pronunciation). Whether these comments are accurate reflections of AAC speech in production is uncertain; however, they do represent linguistic features that are ideologically linked to AAC identity and may be part of perceptual expectations of AAC speech. The fact that these features have been overtly discussed suggests a degree of awareness, possibly due to associations with other socially marked speech styles (e.g., "vocal fry" may available to discuss in AAC speech because of its role in discourse of "Valley Girl" styles). This may make them especially cognitively relevant, or salient, in perceptual processing. At the same time, availability for discussion is only a clue for potential salience as certain features, including segmental ones, may be less likely mentioned because of a lack of folk terminology rather than a lack of awareness (see, e.g., Preston, 1996).

Post-hoc auditory and/or acoustic analyses of perceptual stimuli suggest additional features at the segmental level with the potential to be perceptually associated with AAC ethnicity, despite not having been identified in commentary. In P. Wong and Babel (2017), some features uncovered by post-hoc transcription of the most accurately identified Cantonese/Chinese Canadian talker included: less constricted post-vocalic /ɹ/, occasional /ð/-stopping, word-final coronal deletion, and less reduction in unstressed syllables (conceptually linked to prosodic rhythm). The corresponding best-identified East Indian talker's distinct features included: longer, backer, more monophthongal /oʊ/ and high terminal rise. Via a similar method of auditory analysis examining the talkers in their study most likely rated as "Asian", A. Cheng and Cho (2021) report that /ð/-stopping, relatively long VOT or aspiration in word-initial and word-medial /t/, and word-final devoicing may be linked to their perceptual results. They also examine mean pitch and pitch range, finding that mean pitch was associated with ethnicity ratings for women but not men. Higher pitched women's voices were more likely rated as Asian, while lower pitched women's voices were more likely rated as white.

Finally, Newman and Wu (2011) systematically examined phonetic features in the production of all their speech samples, with some comparison to their perceptual results. Some potential links emerged between a selection of suprasegmental and segmental phonetic variables and "Asian" identity. Greater breathiness was a feature identified as descriptively present in many of the (Korean or Chinese) Asian speakers, and its presence appeared to match up with those who were more accurately identified as "Asian". These data also point to rhythmic variability, longer voiceless stop VOT, lower /ɛ/, and lower onset /ɹ/ as possible perceptually-relevant features of "Asian" ethnic identification. However, these results are difficult to extrapolate as identified effects varied across ethnicity and gender and their sample was small (two speakers per gender-ethnicity category). Moreover, the presence of any particular phonetic feature as well as the number of features were inconsistent even among those who were most recognizable as Asian American.

4.2.2 Linguistic Ideologies of AAC Speech Styles

Given the dearth of knowledge on this topic, these previous studies provide highly informative insight into potential perceptual cues to AAC talker identity. The main goal of all these inquiries could be seen as identifying the existence of features linked to the broad categories of race ("Asian American") or ethnicity ("Chinese American") in perception. To expand on the groundwork laid by these studies, two main issues relating to ideology are worth carefully considering as research moves forward in characterizing listeners' knowledge of AAC speech styles.

First is consideration of the degree of awareness or salience tied to different (potential) cues to AAC talker identity. Newman and Wu (2011) suggest the possibility that, based on the featural inconsistencies across talkers perceived as "Asian", multiple "weak" cues may work together to induce impressions of ethnicity rather than a particularly notable or "salient" feature (for a similar discussion of African American ethnic identification, see Thomas & Reaser, 2004). However, while this is certainly a possibility—and as the authors acknowledge—we cannot yet rule out the possible existence and presumed greater influence of features in the speech that are more strongly associated with Asian identity, such as those within listeners' awareness. Not all linguistic variables behave equally; for example, awareness of a variable is proposed to affect its production patterns in the context of sound change (e.g., Baranowski, 2013; Labov et al., 2013). In perceptual studies, awareness of particular stereotyped features have been suggested to effect special influence on speaker identification judgments, such as regional origin (e.g., Clopper & Pisoni, 2006; Montgomery & Moore, 2018). We can also expect salience to vary at the individual level, such that listeners may attend to different features on the whole or at different times. If so, alongside clarifying or identifying potential cues to Asian American speech, characterizing the ideological importance of different linguistic features, both on average and individually, will help to explain AAC perceptual outcomes.

The second relevant issue to account for is how (potential) cues to AAC identity are tied to different subsets of the category. For example, Newman and Wu (2011) found that different phonetic variables in production associated with "Asian" identity were mainly produced by certain groups, like lower rhythmic variability for Chinese men, or longer VOT for Korean talkers. In terms of perceptual impact, A. Cheng and Cho (2021) report that higher pitch appeared to increase "Asian" ratings only for women's voices, not men's voices. This suggests that different linguistic features are likely tied to particular subcategories of AACs, which could be ones like "Korean Americans", "Chinese American men", or "Asian American women". However, along with these broadly intersectional subcategories are other subcategories of social types or personae that listeners have access to, and which elicit different linguistic ideologies and expectations (e.g., MacFarlane & Stuart-Smith, 2012; D'Onofrio, 2018a, 2019). Like those of particular linguistic features, the ideologies that people hold about types of Asian Americans (even about the broader category itself) varies across individuals. Characterizing the different types of AACs that are recognized by listeners, as well as the aspects of speech that bring up different AAC identities, should further contribute to explaining behavior surrounding perception of AAC speech.

Both of these issues intersect with culturally circulating linguistic ideologies surrounding those in North America racialized as Asian. While the concept of Asian American varieties of English appears less strongly enregistered or "distinctive" than others such as African American or Latino varieties, evidence suggests that it does exist to some degree for some people (see, e.g., metacommentary in Hanna, 1997; Newman and Wu, 2011). Still, as Reyes and Lo (2009) observe, the main linguistic associations with AACs are of "foreign" languages, a "foreign accent", and/or "non-native English", such as the exaggerated, stereotyped variety of "Mock Asian" (Chun, 2004). This is closely tied to ideologies of "forever foreigner" (Tuan, 1998), where Asian-racialized individuals are viewed as "non-English speaking foreigners" (Reyes & Lo, 2009) and less American (Devos & Banaji, 2005). These ideologies about "foreign" or "nonnative" accents cannot be easily disentangled from ideas about and perceptions of "local" AAC speech, even if they are not the same. Further complicating this relationship is that some features of local AAC varieties may indeed be shared to some extent with bilingual contact varieties of English, possibly sourced from particular heritage varieties as an identity marker. Local AAC individuals with a greater degree of orientation towards their heritage as part of their identity may make more use of those linguistic features than others in their production (e.g., L. Jeon, 2017), and this may in turn result in a greater likelihood of their ethnic identity being accurately judged by listeners (e.g., Nagy et al., 2020).

Accordingly, one particular social dimension on which AACs may be believed to differ saliently is "local" versus "foreign", or within the local context, "mainstream" versus "ethnic". These beliefs build off of broader societal ideologies of race, leading Asian Americans to divide other Asian Americans into sub-ethnic ideological categories in a process of intraethnic othering (Pyke & Dang, 2003). Specifically, drawing on interviews with second-generation Vietnamese and Korean Americans from California, Pyke and Dang (2003) demonstrate that this othering takes place along a continuum of acculturation represented at one extreme by the "too ethnic" "FOBs" and at the other extreme by the "too assimilated" "whitewashed". Most individuals claimed themselves as part of the "bicultural middle" in between the two extreme ideological constructs, thus positioning themselves as "normals".

The concept of FOB ("fresh off the boat") has been described as representing a stereotyped style of Asian American who is seen as aligned with recent immigrants (though not necessarily being recent immigrants) by being more oriented towards the ethnic or foreign culture in some way, often with negative connotations and linguistic associations (e.g., M. Jeon, 2007; Shankar, 2008). For example, in the context of Korean Americans, second-generation college students depict the FOB persona as identifiable through behavior (e.g., "acting like Koreans with Korean attitudes", "always hang[ing] out with other FOBs"), visual features (e.g. hairstyle), as well as linguistic style (e.g., language choice; M. Jeon, 2007). In a Desi (South Asian American) high schooler context, the exact associations and formulation of what counts as "FOBby" are different, but the core concept remains and is situated within a dichotomy between mainstream "populars" (analogous to Pyke and Dang's "normals") and nonnormative "FOBs" (Shankar, 2008).

Linguistically, the FOB persona is often marked by greater and/or divergent use of the relevant non-English language(s), as well as linked to less fluent English or a "foreign accent". In fact, Pyke and Dang (2003) point out that "accent and foreign language were particularly important markers of the 'FOB'", corresponding to the importance of these linguistic aspects in anti-Asian discrimination (156). Altogether, ideological ethnic subcategories based on ideas of race, ethnic orientation, and foreignness appear to be available to many AACs, and are inextricably tied to ways of speaking.

4.2.3 Complicating Perception of AAC Speech

What does taking into account these ideologies mean for expectations and perceptual processing of AAC speech? Broadly, it suggests that people have multiple ideas about how an Asian-racialized individual may speak, including more local or foreign-sounding English, which can be drawn on during speech processing. In support, M. Babel's (2022) results suggest that listeners in Vancouver, Canada can expect Asian-racialized talkers to produce either mainstream "L1 English" or Mandarin-accented "L2 English", based on experience with both in the local environment. In addition, particular expectations of speech as local- or foreign-sounding may be induced when a speaker is identified as being aligned with a "mainstream" or "ethnic" orientation.

D'Onofrio (2019) provides an example where different visual styles embodied by an East Asian individual—inferred to represent certain personae—led to differences in linguistic recall performance, as well as ratings of various social and linguistic characteristics. The two styles could be interpreted as examples aligning with mainstream- and ethnically-oriented AAC personae: One photo was rated as more likely to be from the US (as opposed to from a country in Asia) and less likely to have a foreign accent compared to the other photo. Importantly, the mainstream-oriented persona elicited similar recall accuracy to the white comparison individual (suggesting a closer alignment in speech expectations of mainstream AAC to white), while the ethnically-oriented personae elicited a different recall pattern (see also Rubin, 1992; M. Babel & Russell, 2015; McGowan, 2015).

There are also consequences for the theorization of sociophonetic representations and processes underlying perceptual ethnic identification of AACs, as well as the goal of identifying perceptual cues to AAC identity. The incorporation of differently-acculturated AAC personae into our understanding of AAC speech means that we need to consider how specific linguistic features are associated not only with "Asian" as a category but also with subcategories of "Asians". The notion that awareness of a sociophonetic association results in more salient and strongly-weighted associations indicates that we must also account for certain features playing a more prominent role. In other words, different linguistic features may hold stronger or weaker associations with a macrosocial "Asian" category and/or with personae categories like "FOB Asian", "Normal Asian", and "Whitewashed Asian", and those with stronger associations should effect greater influence on perceptual behavior. Note that based on broadly circulating racialized ideologies, we might expect a stereotypical "Asian" category to overlap with the "FOB Asian" persona in terms of foreign associations.

Exactly how these elements play out in production and perceptual representations is up for discussion. For instance, any particular feature (e.g., /oʊ/-backing) might be associated with any of the AAC social categories, including associations with only a single persona (e.g., "FOB Asian") or associations with a combination (e.g., "FOB Asian" and "Asian"). As such, some features may be shared, but some may differ between personae, such as a cue only indexing "FOB Asian" but not "Whitewashed Asian". If so, cues based on foreign "Asian" accents may not necessarily be informative of local AAC identity. Another possibility to consider is that while shared features might index multiple personae, they might do so with at different levels of cue strength along an axis of local to foreign (e.g., weakly backed /oʊ/ might be associated with "Normal Asian" while strongly backed /oʊ/ is associated with "FOB Asian"). In sum, variants associated with various macrosocial and microsocial AAC categories may come together to contribute a holistic impression of a talker—of which only one social dimension is race/ethnicity.

Further, certain features may be more salient than others for a wide variety of reasons, one of those potential reasons being metalinguistic commentary. Explicit discussion of features is theorized to create or reinforce awareness for sociophonetic associations and strengthen the stored representations (e.g., Drager & Kirtley, 2016). In this context, one possible scenario could be that societally stereotyped associations of "Asian" result in greater weight placed on variants ideologically represented in foreign varieties of English (including "Mock Asian"), thus exerting greater influence on Asian identification of talkers whether or not those are actually the most prevalent features in local AAC speech. As it stands, more perceptual work needs to be done with consideration of these aspects of ideology to build a nuanced understanding of the mental representations of "Asian" or AAC English speech. This includes knowledge of particular features and their ideological weight which will allow us to better interpret ethnic identification behavior and characterize linguistic expectations.

4.2.4 The Current Study

The high-level question that the current study aims to address is this: What linguistic features are part of listeners' representations of AAC speech styles, and to what extent do they influence perceptual identification of talker ethnicity? To summarize the literature reviewed above, a wide range of phonetic variables have been discussed by participants and researchers alike as potentially associated with AAC identity, but given the sparse scholarship on this topic and variable ethnic identities examined, evidence supporting any particular feature's role in perceptual judgments is still tentative.

Nevertheless, several phonetic features have appeared in both ideological metacommentary and auditory/acoustic analysis of AAC speech across studies: speech rhythm (linked to less syllable reduction), /ð/-stopping, and /oʊ/(-backing). These features also appear in performances of "Mock Asian" (Chun, 2004), which further supports the possibility of their ideological link to AAC identity for North American listeners, though the actual degree of awareness is unknown. Also unknown is whether certain features may have stronger links to certain subcategories or social types of AACs, such as those who are (perceived as) more oriented towards their heritage culture, or those who are seen as foreign.

In the current study, I combine quantitative and qualitative approaches, investigating perceptual social judgments of AAC English as well as exploring ideologies and awareness surrounding AAC speech via metalinguistic commentary. I place a particular focus on investigating the role of the /oʊ/-backing variable. This phonetic variable was shown in Chapter III to be broadly different between AAC and non-AAC speaker productions, but has thus far received only limited mention in previous perceptual and ideological data. Because speakers were all Californian Asian Americans, and because listener ethnic identity and geographic location may be related to the likelihood that listeners have familiarity with Asian American speech and the ideologies of interest, I specifically recruit a sample of Asian American participants from California.

During the main perceptual social evaluation task, respondents listen to naturalistic speech samples and are asked to rate their impressions of the speaker's social characteristics, including ethnic identification ratings. I link these findings to the production results from Chapter III, examining whether the specific variable of /oʊ/-backing identified in a speaker's production is related to quantifiable differences in perceptual impressions of "Asian" race/ethnicity and other potentially-associated social characteristics. In addition, I extend previous research on the topic of "Asian" ethnic identification by (a) assessing the extent to which different potential cues to Asian American speech are in listeners' awareness, and (b) directly asking listeners to describe their ideological conceptions of Asian American social categories and speech styles.

4.3 Methods

4.3.1 Auditory Stimuli

The auditory stimuli for this survey came from the same 28 speakers analyzed in Chapter III: twenty-three who self-identified as "Asian American" and five who did not (2 European American, 2 Latin American, 1 mixed). The goal was to create speech samples with multiple examples of the target /oʊ/ vowel that were representative of that speaker's average /oʊ/-backing (measured as mean F2). [1]

To do this, five short auditory excerpts (between 1–5 seconds) from the previously analyzed YouTube video recordings were extracted per speaker. Tokens of /oʊ/ along with surrounding contexts were examined starting from those with the most average F2 values. Boundaries were placed at zero-crossings corresponding to word boundaries, and if possible (given the short time window) at prosodic and/or syntactic phrase boundaries. Excerpts were excluded if they contained audio quality issues (e.g., background noise, nonspeech sounds), speech disfluencies (e.g., pauses, speech errors), or content that could reveal the talker's identity either as a YouTuber or their demographic background (e.g., mentions of ethnicity or language). In addition, if multiple possible excerpts contained the same target /oʊ/ word or very similar lexical and/or semantic content, only one was included for consideration. From the remaining tokens, five excerpts with an overall mean F2 close to the speaker's overall mean were selected, typically the excerpts that contained the top five most average /oʊ/ F2 values for that speaker. Peak intensity

[1] Stimuli are available from the author on request.

was scaled to 70 dB and excerpts were concatenated into a single speech sample per talker. The total duration of each speech sample was between 9.33 and 14.27 seconds (*M* = 12.15 seconds, *SD* = 1.59). The speech sample's mean /oʊ/ F2 values were highly correlated with the speaker's values in the full production analysis (*r* = .98), supporting the validity of linking talker /oʊ/-backing from the production analysis to the perceptual judgments.

4.3.2 Survey & Procedure

The survey was hosted on Qualtrics. Participants were told that the purpose of the survey was to study "what you can tell about a speaker from their voice alone" and "focuses on what people think about Asian Americans and their speech". Note that this description drew participants attention to Asian American speech and identity. The full survey was composed of four sections: the listening task, ideology survey, social network assessment, and personal demographics survey. The listening task was presented as "Part 1" and the remaining survey questions as "Part 2".

While the questions in Part 2 were the same for every participant, there were three versions of Part 1, each containing (a) a subset of AAC talkers and (b) the full set of non-AAC talkers. This was done in order to reduce the length of the survey and even out the ratio of AAC and non-AAC talkers per listener. In addition, to introduce the range of /oʊ/-backing to listeners and maintain an even number of talkers per listener, three talkers were selected to be presented at the start of the survey: the non-AAC talker with the most fronted /oʊ/ (NO1) and the two AAC talkers with the most backed/oʊ/ (AC3, AK4). In the end, each version contained 14 speech samples: the three extreme-/oʊ/ talkers, the four remaining non-AAC talkers, and seven of the remaining 21 AAC talkers. The extreme-/oʊ/ talkers were always randomly presented within the first three trials, and the rest of the talkers were randomly presented across trials four through fourteen.

After consenting to participate in the study, participants first completed Part 1, the listening task, consisting of 14 trials. On each page, they were presented with an audio speech sample along with various questions asking for impressions of the speaker's characteristics. They could play the recording as many times as they wished. The most relevant questions are detailed below.

Based on the speech they heard, participants were asked to rate the degree or likelihood of 13 perceived characteristics on slider scales. The first set of slider scales asked for the extent to which the speaker elicited particular social impressions (0 = *Not at all*, 50 = *Neutral*, 100 = *Extremely*). The six attributes (FRIENDLY, EDUCATED, SHY, CLEAR, NATIVE, ACCENTED) were preceded by "This speaker sounds..." (e.g., "This speaker sounds FRIENDLY", "This speaker sounds like a NATIVE SPEAKER of English").

The next two sets of slider scales asked for likelihood ratings (0 = *No, extremely unlikely*, 50 = *Neutral*, 100 = *Yes, extremely likely*). The first of these asked about regional origin ("Does this speaker sound like they grew up...") and included scales for "Outside of the U.S. or Canada" (INTERNATIONAL), "Somewhere in the U.S. or Canada" (NORTH AMERICAN), and "Somewhere in California" (CALIFORNIAN). The other asked about racioethnic identity ("Does this speaker sound like their ethnic background is..."), including scales for "Asian American" (ASIAN), "Black/ African American" (BLACK), "Latin/Hispanic American" (LATIN), and "White/ European American" (WHITE). If the participant rated the talker over a value of 50 on the ASIAN scale, they were further asked which specific ethnicity they might be from a list of choices (e.g., "Chinese American", "Filipino American") that allowed multiple selections and included an option for "Other" (for which they could type in a different label) and "Not sure". They were also asked to list aspects of the speech sample that stood out, if any. After all the trials were completed, participants were asked what they listened for to guess ethnicity, whether they recognized any of the voices, and whether they noticed anything else about the stimuli or tasks.

In Part 2, the ideology survey began with both open-ended (i.e., short answer) and closed-ended (i.e., multiple choice, 7-point Likert scale) questions about participants' ideas about Asian Americans, different Asian American ethnic groups, and "types" or "stereotypes" of Asian Americans. They were then asked about speech styles associated with each of these categories of Asian Americans. To access more detailed sociolinguistic associations, respondents were next provided with a list of 14 statements describing aspects of speech (linguistic features). The features and descriptions (see Table 4.5) were selected and phrased based on results from previous literature on this topic (e.g., Hanna, 1997; A. Cheng & Cho, 2021), along with a few additional possibilities based on my own impressionistic auditory judgments from listening to the stimuli. For each of these statements, participants were asked to rate on a 7-point Likert scale (a) how familiar it was (1 = *not familiar at all* to 7 = *extremely familiar*) and (b) how much they agreed with it (1 = *strongly disagree* to 7 = *strongly agree*) as a descriptor of AAC speech. Specifically, the question wording was "How **familiar** are you with the following comments that people make about an 'Asian American accent' or 'sounding Asian American', whether you agree or not?" and "In general, how much do you **agree** with the following comments that people make about an 'Asian American accent' or 'sounding Asian American'?"

The final two sections collected information about the background of the participant for context in interpreting results. The social network assessment asked about the percentage of their social network (e.g., family, friends, colleagues) that was Asian in order to estimate participants' familiarity with Asian Americans and Asian American speech. The personal demographics survey asked for information such as age, gender, ethnicity, and locations lived. In total, the full survey took approximately 50 minutes to complete. Each participant who completed the study in full received \$10 in compensation.

4.3.3 Participants

Sixty participants were recruited online via Prolific (www.prolific.com), a participant recruitment website, to participate in the survey remotely using their own devices. All reported using headphones (38 in-ear, 22 over-ear) attached to a desktop or laptop computer (i.e., no participants used mobile devices like phones or tablets). Twenty listeners participated in each version of the survey.

All participants were living in California at the time of the survey; the majority were born and raised there, but those who were not had lived there for at least three years. All self-identified their racioethnic background as “Asian” (one also identified as “Biracial or Multiracial”). A range of specific ethnicities were represented across participants (see Table 4.1 for self-identified ethnicity). The most common ethnicities were Vietnamese, Chinese, and Filipino. In addition, a bal-

Ethnicity	**Total (n = 60)**	**Male (n = 30)**	**Female (n = 28)**	**NB (n = 2)**
Vietnamese	**17**	11	6	0
Chinese	**16**	7	8	1
Filipino	**7**	4	2	1
Japanese	**4**	4	0	0
Indian	**4**	1	3	0
Korean	**3**	2	1	0
Taiwanese	**3**	1	2	0
Chinese-Filipino; Filipino-Chinese	**2**	0	2	0
Filipino-Mexican-Spanish	**1**	0	1	0
Laotian	**1**	0	1	0
Mien	**1**	0	1	0
Pakistani	**1**	0	1	0

Table 4.1: Summary of the number of participants by self-identified ethnicity and gender.

anced gender distribution was attempted during recruitment in order to gather responses from individuals with a range of experiences. Of the 60 participants, 30 self-identified as "male", 28 as "female", and two as "nonbinary".

4.3.4 Data Analysis

4.3.4.1 Perceptual Behavior

Participants' speaker evaluation responses to the perceptual rating task were analyzed as follows. To be able to compare the talkers that appeared as constants across all three versions of the survey (and therefore presented to all 60 participants) to those talkers that only presented in one version of the survey (and therefore presented only to a subset of 20 participants), a random sample of 20 participants was selected to represent the rating data for the seven constant talkers. All responses for the remaining 21 talkers (20 per talker) were then combined with the randomly sampled constant-talker responses (20 per talker) to form the analysis dataset. Three talker trials were removed due to the listener recognizing the talker's voice, leaving 557 trials (each including 13 ratings) to be analyzed.

To assess the relationship between the presence of /oʊ/-backing in a speech sample and speaker evaluation ratings—with particular interest in ASIAN likelihood ratings—mixed-effects linear regression models were fit to the slider scale rating data using *lme4* (Bates et al., 2015) in R (R Core Team, 2020). Separate models were run for each of the 13 social characteristic scales (see list of scales in Table 4.2). The dependent variable for each model was slider scale rating. The fixed effect predictor was TALKER MEAN F2 value for /oʊ/, and random effects included by-participant random intercepts and slopes. The Bonferroni correction for multiple comparisons was applied to interpret the results across the 13 social chracteristic scale models.

4.3.4.2 Ideological Self-Report

The ideological self-report data consisted of all 60 participants' responses to the survey. Descriptive statistics are provided for select 7-point Likert scale ratings asking about ideologies of Asian American speech both generally and for specific linguistic features. To probe in more detail the degree of awareness for different potential linguistic features of an "Asian American accent"—with particular interest in /oʊ/(-backing)—I analyze FAMILIARITY and ACCURACY Likert scale ratings for the 14 linguistic features (see list of features in Table 4.5). Here, ratings of FAMILIARITY are taken represent the extent to which each statement is stereotyped in the respondents' experiences while ratings of agreement are interpreted as the perceived ACCURACY of each statement within the respondents' own beliefs about Asian American speech. In total, 840 responses for FAMILIARITY and 838 responses for ACCURACY were analyzed (2 missing responses for ACCURACY were removed).

Because interpretation of the two Likert scales differed from each other, each was scaled differently. For FAMILIARITY, ratings were scaled such that zero represented "not familiar at all" (i.e., range: 0–6). For ACCURACY, ratings were scaled such that zero represented "neither agree nor disagree" (i.e., range: -3–3). Then, two mixed-effects linear regression models were fit to the Likert scale rating data using *lme4* (Bates et al., 2015) in R (R Core Team, 2020), one for FAMILIARITY and one for ACCURACY. Models included scaled rating as the dependent variable, linguistic FEATURE as the fixed effect predictor, and by-participant random intercepts.To facilitate relative comparison between levels of FEATURE, the reference level for each model was set to the feature with the lowest mean rating out of the 14 statements: For FAMILIARITY, this was FRY (M = 3.07) and for ACCURACY, this was PAUSE (M = 3.65). Using the *emmeans* package (Lenth, 2023), estimated marginal means were extracted for each FEATURE and all pairwise comparisons were conducted; p values were adjusted using the Tukey method.

Secondarily, I explore ideas about different AAC social types, with the goal of identifying the presence of differently-acculturated Asian Americans in respondents' ideologies. Open-ended text data were collected about Asian American social types from two questions: one asking participants to describe any "types" of Asian Americans they were familiar with, and the second asking if any "types" were tied to specific speech styles. For the social category question, text content of all responses were qualitatively coded for common types of Asian Americans in an inductive manner, guided by labels or descriptors used by participants, across multiple rounds of coding. The first round used mainly In Vivo coding (using participants' own words) to extract codes for observations of types. The second round of coding consolidated the various text exemplars and initial codes into unified codes for quantification. The third round finalized codes used for each exemplar, making decisions on edge cases and removing cases where a code was used only once. In the final coding scheme, specific texts could be associated with multiple codes. Coding for the speech style question followed a similar procedure with two differences: (a) because there was less content, the first and second rounds of coding were collapsed into one, and (b) both social types and linguistic features were coded simultaneously, using the social type codes that emerged from the previous question.

4.3.5 Predictions

In the auditory rating task, I expect that listeners on the whole will be able to identify AAC talkers as likely to be "Asian" (i.e., likelihood ratings above 50), and that some talkers will be more likely to be identified as ASIAN than others. I predict that the presence of greater /oʊ/-backing will be related to an increase in ASIAN likelihood rating, whether due in part to the role of /oʊ/-backing as a cue of AAC identity or co-occurring with other AAC-linked features. With a more exploratory focus, I expect that /oʊ/-backing may also be related to ratings for social characteristics relating to Asian-racialized stereotypes, such as being foreign (INTERNATIONAL).

In the ideological commentary survey, I expect that, when rating subjective familiarity and accuracy of the association between various linguistic features and "sounding Asian", respondents will express some degree of awareness (i.e., ratings above 1 for FAMILIARITY, ratings above 4 for ACCURACY) for /oʊ/(-backing) on average. At the individual level, awareness is expected for some respondents and not others. At the same time, I do not expect /oʊ/ to be the most salient feature. Finally, I expect participants to show recognition of Asian American social types along the axis of mainstream-oriented to ethnically-oriented (including the concept of "FOBs"), and to ideologically link these types to speech styles to some degree.

4.4 Results: Perceptual Behavior

In this section, I report on participants' social evaluation responses to the auditory rating task that involved perceptual ethnic identification of both AAC and non-AAC talkers.

4.4.1 By-Group Ratings

The first broad question under investigation is how Asian American talkers may be rated across various social dimensions, including perceived racioethnic background. Table 4.2 summarizes the mean values per rating scale by AAC status as well as across the full sample (i.e., overall). Figure 4.1 visualizes the mean rating values for each survey rating question by talker category (AAC or non-AAC).

I first consider mean ratings across the full sample—which, recall, is heavily skewed towards Asian American talkers—to provide a birds-eye view of this data.

Rating Scale	**AAC (*n* = 23)**	**Non-AAC (*n* = 5)**	**Overall (*n* = 28)**
Accented	29.43 (9.73)	20.42 (16.65)	27.82 (11.43)
Asian	67.64 (7.95)	50.55 (12.26)	64.59 (10.87)
Black	16.75 (4.23)	21.62 (6.58)	17.62 (4.96)
Californian	60.47 (6.65)	70.28 (5.11)	62.22 (7.38)
Clear	76.28 (5.64)	77.86 (7.73)	76.56 (5.92)
Educated	69.4 (4.34)	60.21 (3.95)	67.76 (5.52)
Friendly	68.5 (4.4)	68.11 (4.87)	68.43 (4.39)
International	28.08 (8.46)	21.36 (14.89)	26.88 (9.9)
Latin	23.6 (5.36)	34.75 (20.2)	25.59 (10.14)
Native	79.27 (7.99)	85.1 (10.55)	80.31 (8.59)
North American	78.03 (6.11)	86.17 (9.2)	79.48 (7.28)
Shy	32.99 (10.98)	14.7 (6.35)	29.72 (12.45)
White	38.37 (10.99)	55.5 (20.49)	41.43 (14.33)

Table 4.2: Mean ratings *(SD)* per rating scale for AAC talkers, non-AAC talkers, and the overall average across all talkers. Shading indicates overall values > 50.

Overall, listeners perceived talkers as likely to be NATIVE, NORTH AMERICAN, and CALIFORNIAN, as well as EDUCATED, FRIENDLY, and CLEAR. Conversely, talkers on the whole were rated as unlikely to be ACCENTED, INTERNATIONAL, and SHY. These results are conceptually consistent with the particular speech samples used: Talkers did indeed grow up speaking English in North America, specifically California, and, as the clips were pulled from YouTube videos, it is reasonable to expect many of them to speak in a clear, friendly, and outgoing manner.

In terms of race/ethnicity, this sample of talkers was rated as most likely to be ASIAN, followed by WHITE, LATIN, and BLACK. This again broadly aligns with the actual sample demographics, although the likelihood of WHITE ratings are much higher than the actual proportion of European American talkers. This may be attributed to some Asian American speakers sounding to listeners as likely to be white and/or an effect where white is the "default" racioethnic expectation in the absence of defining linguistic features (e.g., P. Wong & Babel, 2017; A. Cheng & Cho, 2021).

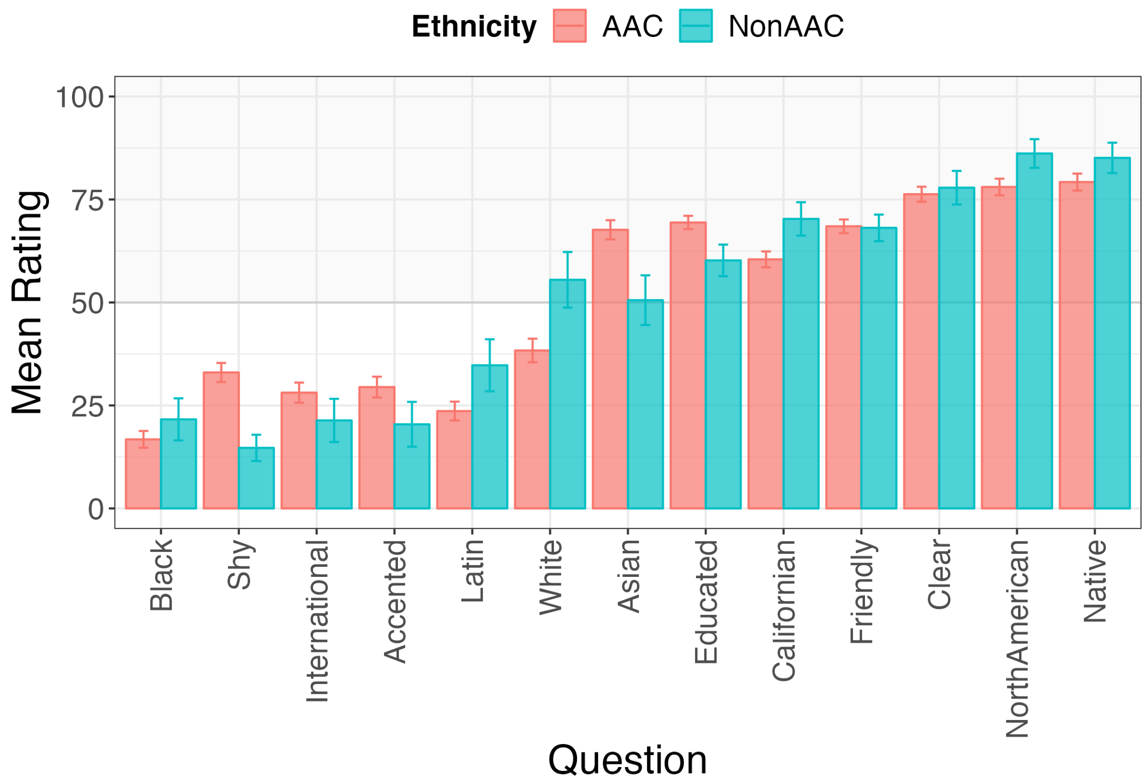

Figure 4.1: Mean ratings (error bars represent 95% confidence intervals) per social characteristic slider scale, representing listener impressions of each talker based on their speech samples. Values are averaged across listener and talker, where talker is grouped by AAC (n = 23) and non-AAC (n = 5) status.

To probe the question of whether listeners accurately identified Asian American speakers on the whole, I turn to average ratings by AAC status where I interpret rating values over 50 to indicate a decision of "likely to be Asian". As Figure 4.1 shows, Asian American talkers were on average rated likely to be ASIAN (M = 67.64) and unlikely to be WHITE (M = 38.37) or LATIN (M = 23.60), suggesting that participants were able to use aspects of speech to guess the likely ethnicity of at least some Asian American talkers. In comparison, the small sample of non-Asian American talkers (including both European and Latin Americans) were socially perceived as less likely to be as ASIAN (M = 50.55) and more likely to be WHITE (M = 55.50) or LATIN (M = 34.75). In terms of other social characteristics, Asian Americans, as compared to non-Asian Americans, were numerically rated less likely to be NORTH AMERICAN, NATIVE, and CALIFORNIAN as well as more likely to be ACCENTED, SHY, INTERNATIONAL, and EDUCATED.

Overall, then, it appears that listeners were able to consistently identify some of the Asian American talkers as "Asian", and that this impression may be linked to talkers sounding less likely to have grown up locally in California or the U.S. Nevertheless, all talkers were generally rated as sounding quite likely to be both North American and native speakers of English (with mean ratings around 78-80). This suggests that listeners generally believed they were specifically identifying Asian Americans—talkers of Asian ethnic background in the local context—not foreign or international Asian individuals.

4.4.2 By-Talker ASIAN Ratings

Examination of perceptual judgments for each individual talker help to shed light on which talkers were accurately identified and, in conjunction with the previous exploratory cluster analysis of each talkers' production, what listeners may have been attending to in order to make ethnicity judgments. These data are supplemented by listeners' explicit commentary about what aspects of speech stood out to them, providing further hints about the cues that may underlie response patterns. Figure 4.2 presents the distribution of ASIAN ratings per talker. To add to this, Figure 4.3 displays by-talker mean ASIAN ratings where talker is colored either by their ethnicity category or their assigned cluster based on speech production patterns (see 8-cluster solution results reported in Section 3.4.4).

Which talkers did listeners identify as most likely to be Asian American? Figure 4.2 reveals that while some talkers elicited more variability in perceptual impressions of ethnicity, certain AAC talkers were always or nearly always considered likely to be ASIAN (e.g., AK4, AC3, AV2, AK3, AS2). In fact, all listeners rated AK4

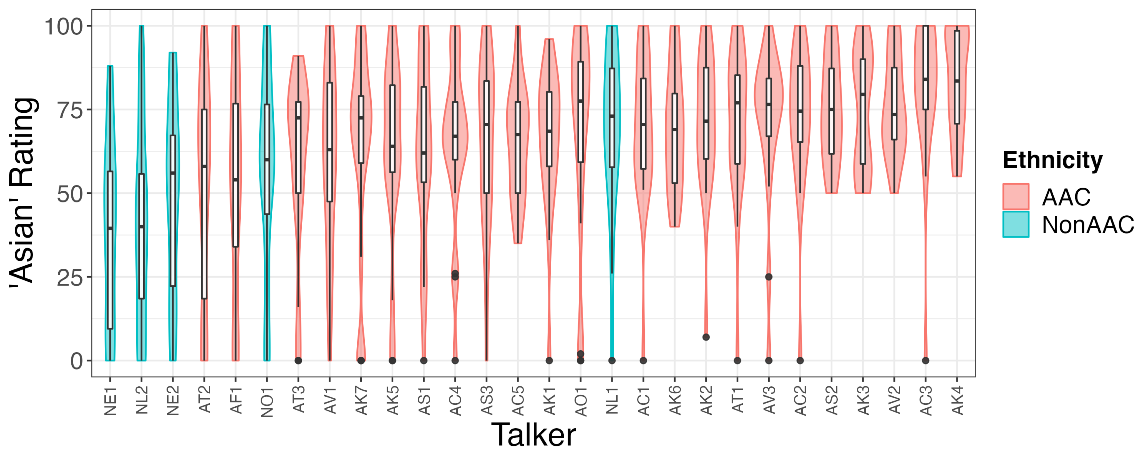

Figure 4.2: Violin and boxplots showing the distribution of ASIAN rating scores by talker.

above 50 (*neutral*) on the ASIAN scale, and all but one rated AC3 and AV2 above 50 as well. Non-AAC talkers were rated as overall less likely to be ASIAN than AAC talkers, and we see that in the individual talker data as well—especially for the two European Americans (NE1, NE2) and one of the Latin Americans (NL2). This together indicates that there is some consistency in listener judgments of Asian American identity that align with actual "Asian" identity.

Next, based on visual inspection of Figure 4.3, no particular pattern by ethnicity is present. That is, no ethnic groups were clearly more likely to be identified as "Asian" than others. Related to this finding is that listeners did not seem able to consistently label the specific ethnicity of speakers that they identified as racioethnically Asian. For example, the most commonly selected ethnic group for the top three Asian-perceived speakers (AK4, AC3, AV2) was Chinese American, which was correct in only one case. Some speakers, both Korean and non-Korean, were considered most likely Korean American (e.g., AK2, AC1), while other speakers appeared to be unidentifiably Asian (e.g., the top choice selected for AC2, AS1, and AC4 was "not sure"). Instead of response patterns based on ethnicity, descriptive patterns emerge in the ASIAN rating responses based on clusters of talkers

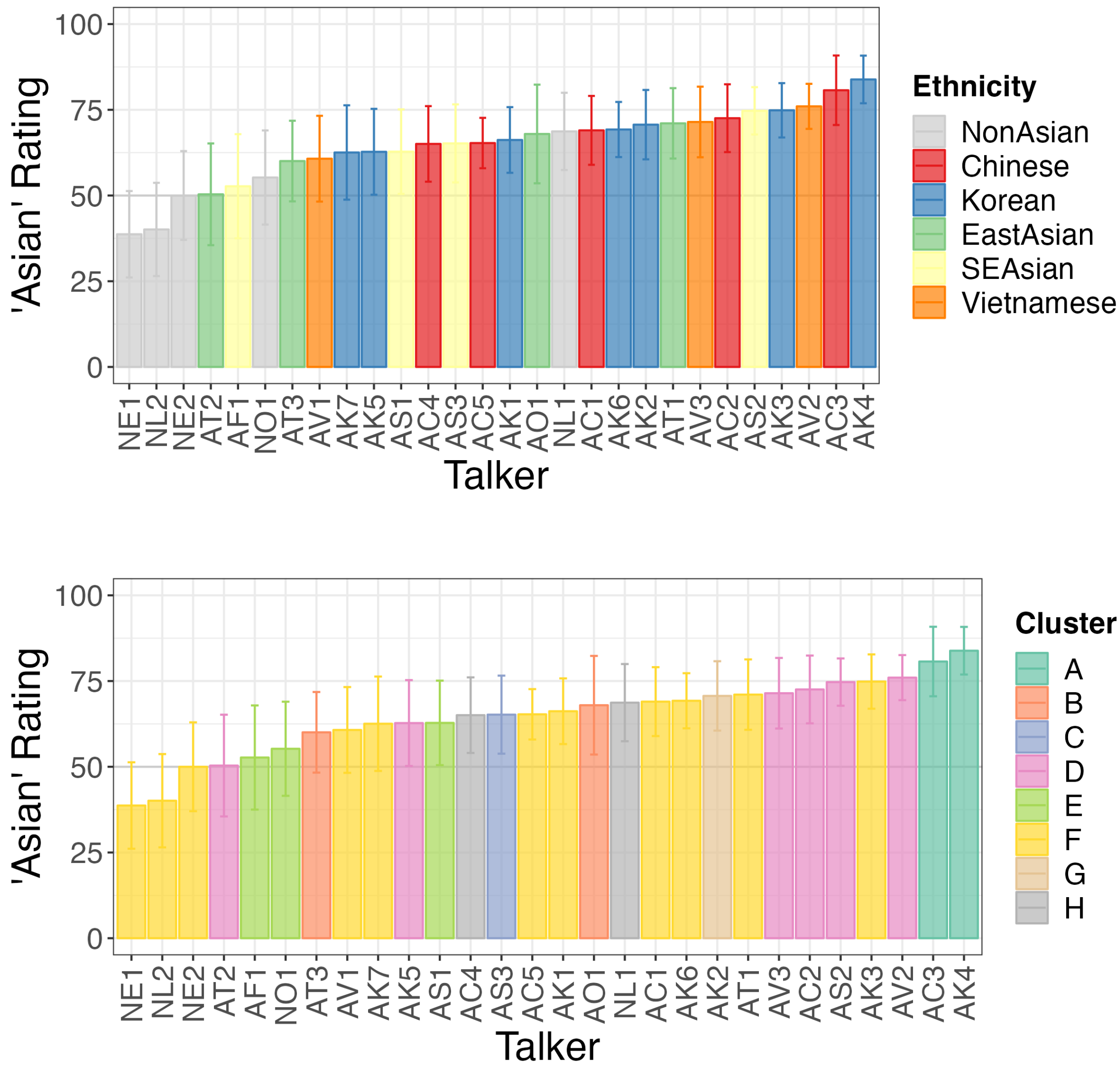

Figure 4.3: Mean ASIAN ratings by talker (error bars represent 95% confidence intervals), colored by (top) ethnicity grouping or (bottom) assigned cluster in the 8-cluster solution from prior clustering analysis.

with distinctive phonetic features in production.

First, the two Cluster A talkers (AC3, AK4) with extremely backed /oʊ/ were noticeably rated as most likely to be ASIAN. Similarly, the single talker from Cluster G (AK2) who had both relatively backed and monophthongal /oʊ/ was rated as relatively likely to be ASIAN. These patterns are consistent with /oʊ/-backing being used as a perceptual cue. Indeed, /oʊ/ or words with /oʊ/ were mentioned in at least one comment for all of these speakers. For AK4 (n = 2), one respondent highlighted the "[a]ccent and intonation on 'know'" as notable to them and the other suggested that the speaker "round[s] out their 'o' sounds". AC3 (n = 2) received a comment that "the pronunciations of the words 'clothes' and 'growth' made me think the speaker was [A]sian" while another described their "o" sounds as "somewhat elongated". Finally, AK1 (n = 1) elicited a detailed comment expressing that it was "[r]eally hard to tell" but "the word 'know' gives more of a [T]aiwanese [A]merican sound since it kind of lags around the middle of the word." Although it is not transparent how non-technical descriptions of /oʊ/ correspond to specific phonetic detail, descriptions of roundedness and length may offer some hints about the relevance of backing or monophthongization. Moreover, this commentary also explicitly ties the vowel and its properties to an "Asian" or specific ethnic category (even though the association with "Taiwanese American" did not happen to be accurate in this case), strongly implying that /oʊ/ contributed to "Asian" impressions for some listeners at a conscious level.

Of additional interest are the mentions of another feature, /ð/-stopping, that co-occurred in the speech of the top-ranking Cluster A talkers as well. These occurred at a similar or slightly higher rate than /oʊ/. Specifically, AK4 elicited three similar comments pointing out that the "speaker pronounced 'them' with a 'd' sound rather than a 'th'". For AC3, three respondents referenced the same phrase of "all over the place", with one directing focus to the word "the" and another directly highlighting "their pronunciation of 'the' as 'da'" as the salient feature. This may suggest that, along with /oʊ/, it was a feature available in consciousness that may have influenced judgments of "Asian" likelihood.

Coming back to the observation that strong /oʊ/-backing coincided with high ASIAN ratings, the corresponding opposite pattern appears to exist as well: Cluster E (AF1, NO1, AS1) had particularly fronted /oʊ/ (along with lower nPVI) and were relatively unlikely to be rated as ASIAN. As the talker with the most extreme /oʊ/-fronting, AS1 elicited comments about their "stressing of the word 'so'" and their production of "so gross" being "very Californian", a "valley girl phras[e]", and "almost European sounding"; these are somewhat ambiguous as to the relevant feature under discussion but could have been linking the obviously fronted /oʊ/ to a Californian, valley girl, and possibly white identity. For the two lower-rated speakers, no explicit commentary mentioned /oʊ/, which may be reflective of relative /oʊ/-fronting as the local mainstream. However, potentially related is that sounding "Californian" was specifically mentioned for NO1 three times and "valley girl" was mentioned once, with one respondent suggesting that "vowels and word choice" were the deciding factors for this impression (use of "like" was also mentioned thrice). Together, these results align with expectations that /oʊ/-fronting is not associated with "Asian" (though other features of these speakers' speech may have been), and instead appears more aligned with Californianness and whiteness.

For additional context, many of Cluster D's talkers—those who had particularly diphthongal /oʊ/ paired with somewhat backed /oʊ/—were rated as quite likely to be ASIAN (namely, AV2, AS2, AC2, and AV3 but not AT2 or AK5). Finally, a number of talkers from Cluster F—those with relatively monophthongal /oʊ/ and roughly average /oʊ/-backing—were rated as relatively likely to be ASIAN (especially AK3 and AT1) but many of them also were not; note that this cluster was rather large and includes the three non-AAC talkers rated least likely to be ASIAN along with various AAC talkers. Some of the higher-ranking speakers of these clusters elicited comments about /oʊ/ in some form, but not all did (e.g., AV2 but not AS2), which may speak to the variety of other cues available for use in the ethnic identification task. Of the full set of comments, only a small percentage (n = 27, 6%) mentioned /oʊ/ or a word containing /oʊ/ as salient (note that some of these comments may not be referring to the pronunciation of /oʊ/ at all), and only a subset (n = 12, 3%) of these were descriptive enough to be interpreted as likely referring to /oʊ/ or the pronunciation of /oʊ/ specifically.

In all, this individual-level data provides useful insights into possible perceptual features of Asian American identity. The main observation is that /oʊ/-backing seems to be a relevant feature used by at least some listeners to perceptually identify "Asian" speakers. Talkers from clusters identified as having more retracted /oʊ/ were often rated as more likely to be "Asian" and vice versa. Explicit commentary supports both of these observations, further suggesting not only that /oʊ/(-backing) was tied specifically to "Asian" identity in awareness for some people, but also that /ð/-stopping appeared to be a relevant and consciously-accessible perceptual cue to "Asian" identity as well.

Given that the stimuli were constructed with a focus on only selecting representative /oʊ/-backing tokens per speaker, we cannot make reasonable conclusions about the other two features' role in this study. However, the mixed results for other clusters and the large variety of open-ended comments suggest that other linguistic features—monophthongal /oʊ/, speech rhythm, and/or various other features that were not measured in the production study—likely played a considerable role in influencing Asian ethnic identification responses in this task.

4.4.3 /oʊ/-Backing By-Talker Ratings

To more directly examine the question of how /oʊ/-backing (and other likely-present associated linguistic features) may have contributed to social impressions of these speakers, I next explore the link between each perceived characteristic with talker /oʊ/-backing. Figure 4.4 shows the relationship between selected perception rating means per talker and that talker's mean /oʊ/ F2 value. The x-axis is reversed such that lower values (representing more backed /oʊ/ production) are presented towards the right side of each graph.

Looking at Figure 4.4, these data appear to support relationships between /oʊ/-backing (along any with co-occurring linguistic features) and selected perceptual impressions of a speaker's voice. Specifically, talkers who produced more /oʊ/-backing appear to be rated as more likely to be ASIAN, INTERNATIONAL, and ACCENTED while also being perceived as less likely WHITE, NORTH AMERICAN, and NATIVE speakers of English. To a lesser extent, visual inspection suggests there may also be a relationship of /oʊ/-backing with higher likelihood of being EDUCATED and lower likelihood of being SHY.

The relevant model outputs are summarized in Table 4.3. To compare across

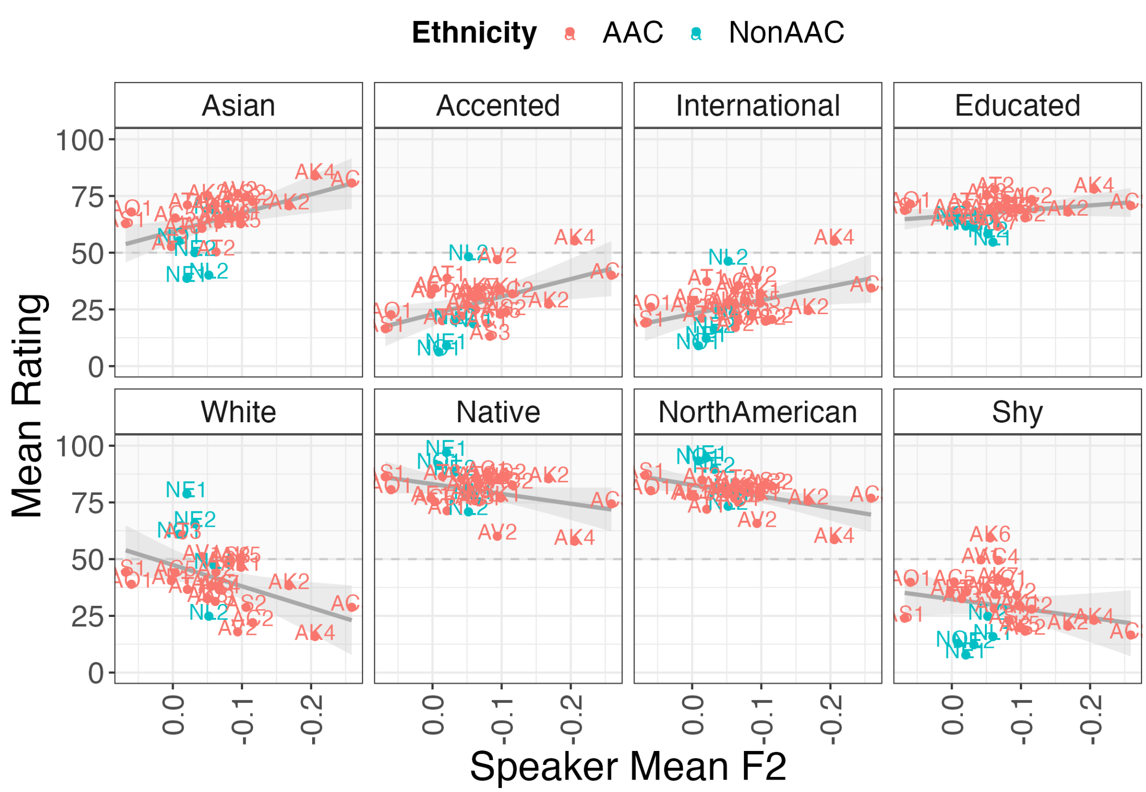

Figure 4.4: Mean ratings (0-100) for selected sliding scale questions from the auditory rating task, averaged across participant and talker. Talker is colored by AAC (n = 23) and non-AAC (n = 5) status. The x-axis is reversed such that lower F2 values, representing more backed vowels, are located to the right.

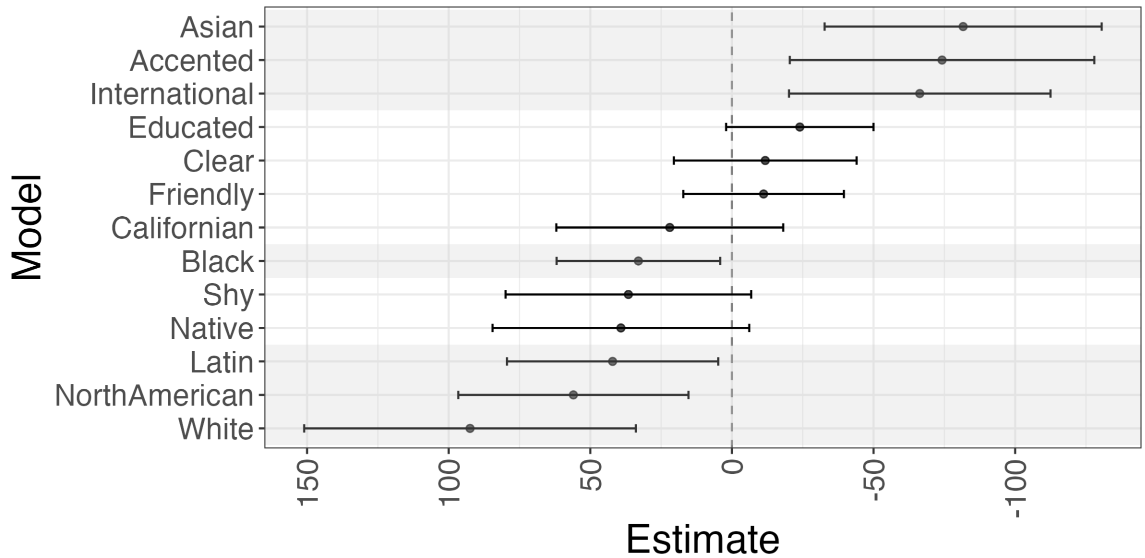

Figure 4.5: Model beta coefficient estimates (error bars represent 99.6% confidence intervals) of the predicted effect of talker mean /oʊ/ F2 for each perceived talker rating model. Models with confidence intervals excluding zero (i.e., a p-value $<$ .0038) are highlighted in grey. The x-axis is reversed to align with typical formant visualizations where lower F2 values, representing more backed vowels, are located to the right.

Model	β	***SE***	***t***	**CI lower**	**CI upper**	***p***
Asian	-81.630	16.075	-5.078	-130.541	-32.719	< .001
Accented	-74.188	17.890	-4.147	-127.947	-20.429	< .001
International	-66.318	15.202	-4.362	-112.484	-20.153	< .001
Educated	-23.968	8.974	-2.671	-49.985	2.049	.008
Clear	-11.765	10.575	-1.112	-44.026	20.497	.272
Friendly	-11.182	9.216	-1.213	-39.536	17.172	.233
Black	33.022	9.940	3.322	4.147	61.897	.001
Shy	36.553	14.252	2.565	-6.803	79.910	.014
Native	39.196	15.153	2.587	-6.121	84.512	.012
Latin	42.131	12.886	3.269	4.839	79.423	.001
Californian	21.943	13.036	1.683	-18.113	61.999	.102
NorthAmerican	55.974	13.513	4.142	15.336	96.612	< .001
White	92.461	19.152	4.828	33.907	151.014	< .001

Table 4.3: Model output summaries for the predicted effect of talker mean /oʊ/ F2 for each perceived talker rating model, ordered by beta coefficient estimates. Models with 99.6% confidence intervals excluding zero (i.e., p-value $< .0038$) are highlighted in grey.

models, the estimates (slopes) of perceived talker rating by F2 are visualized in Figure 4.5 (the x-axis is reversed to be comparable to the typical reversed F2 axis in previous graphs). These results confirm that talkers who produced lower /oʊ/ F2 values (i.e., more backed /oʊ/) were reliably rated as more likely to be ASIAN, INTERNATIONAL, and ACCENTED as well as less likely to be WHITE and NORTH AMERICAN. Lower /oʊ/ F2 values in the speech sample also predicted lower likelihood ratings of LATIN and BLACK racioethnic identity. No other social impression ratings were clearly found to pattern with /oʊ/ F2 measures, though there were trends of lower F2 being associated with increased EDUCATED ratings along with decreased NATIVE and SHY ratings.

To summarize, this outcome indicates that hearing speech samples with a greater degree of /oʊ/-backing led listeners to interpret the speaker as likely to be Asian but not white, Latin, or Black; that is, the presence of /oʊ/-backing was clearly associated with Asian-specific impressions in this study. This effect was paired with impressions of greater accentedness as well as an increased expectation that the speaker was specifically an international or foreign (Asian) speaker rather than one from North America. Such a result seems to conceptually link together Asianness with accentedness and foreignness.

4.4.4 Interim Discussion

The perceptual rating data (a) shows that self-identified AAC speakers can often be identified as racioethnically Asian, and (b) supports the possibility that /oʊ/-backing may be a linguistic feature that listeners use to make racioethnic judgments about speakers, particularly about whether speakers are Asian or not (mainly in opposition to whiteness). Of course, due to the design of this study, which uses naturalistic speech samples, /oʊ/-backing cannot be isolated from other linguistic features present in the auditory signal that listeners may be attending to and using to inform their ethnic identification responses. That is, listeners may not in fact be attending to /oʊ/-backing but instead picking up on co-occurring features that index Asian speaker identity. Alternatively, /oʊ/-backing may be a useful cue but only in conjunction with other co-occurring cues. While these results cannot confirm that /oʊ/-backing was actively being used by listeners to make judgments, the evidence shows that /oʊ/-backing is at least passively present in the signal, presumably together with other indicators of Asian identity, when speech is perceived as "sounding Asian". Still, possible support for the active use of /oʊ/-backing comes from listeners' metalinguistic commentary provided after making each evaluation, where the pronunciation of /oʊ/ was consciously identified by some listeners as a salient feature of the signal, along with other features like /ð/-stopping. In the next section, I add on to these findings by specifically investigating participants' social awareness of /oʊ/-backing relative to other possible linguistic features to further understand whether and to what extent /oʊ/-backing may be ideologically linked to Asian (American) speech.

4.5 Results: Ideological Self-Report

In this section, I detail the outcomes of the survey questions probing ideologies about AACs, including beliefs about linguistic features of Asian American speech and social types of Asian Americans. On the whole, responses for all questions spanned the entire range of possibilities, suggesting that there were wide varieties of opinions and experiences with ideologies of Asian Americans and their speech.

4.5.1 Asian American Speech

To start, I review participants' ideologies about Asian American speech at a general level. In these questions, "Asian American accent" and "sounding Asian

(American)" refer to the same concept. Figure 4.6 displays the selection distributions for 7-point Likert scale questions asking respondents for subjective ratings of (a) their familiarity with the idea of "sounding Asian (American)", (b) the quantity of people in their experience who talk about the idea of "sounding Asian (American)", and (c) their accuracy rate for identifying someone as Asian American when only hearing their voice.

In general, a majority of participants felt that there were or could be differences between how Asian Americans and other Americans speak ("yes" = 67%, "it depends" = 12% , "no" = 22%). When asking directly about the idea of "an Asian American accent" or "sounding Asian American" via Likert scale (1 = *not familiar at all* to 7 = *extremely familiar*), this sample reported that it was a familiar concept (M = 4.68, *Median* = 5, SD = 1.82). The distribution visualized in Figure 4.6 highlights the high variability and left skew. Only 10% selected 1 (*not familiar at all*). In other words, 90% of participants were at least slightly familiar with "sounding Asian".

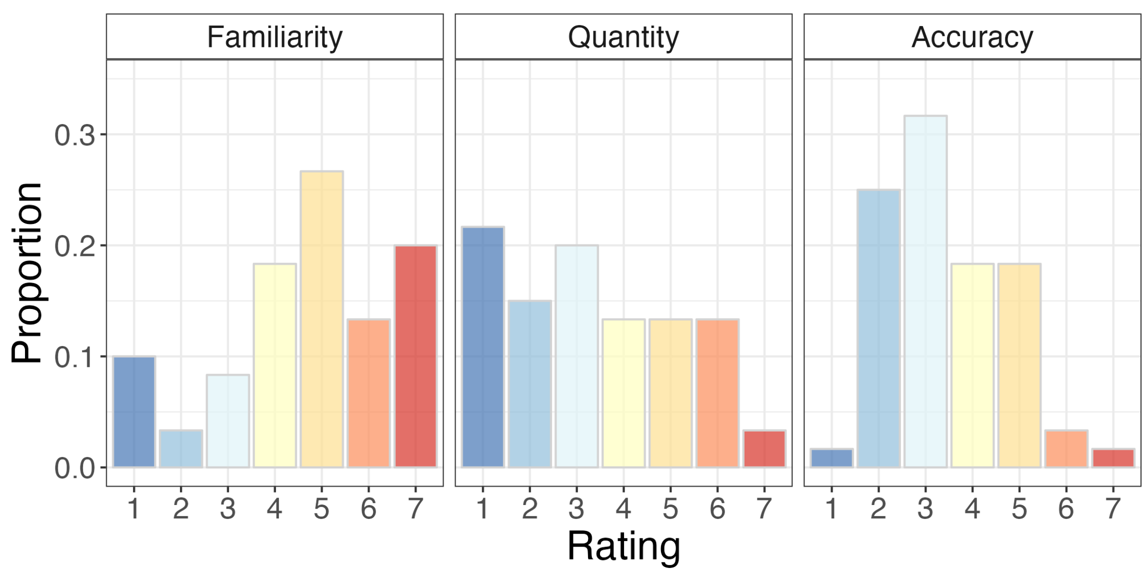

Figure 4.6: Proportion of participants who selected each point on 7-point Likert scales for subjective ratings of (a) familiarity with "Asian American accent" as an idea, (b) quantity of people who talk about an "Asian American accent", as well as (c) accuracy rate for identifying Asian American speakers.

The most common response was 5 (*quite familiar*; 27%), but a sizable portion of participants also selected 7 (*extremely familiar*; 20%).

In terms of exposure to the idea through metalinguistic commentary, respondents indicated on a Likert scale from 1 (*nobody*) to 7 (*everyone*) that some people do talk about an "Asian American accent" (M = 3.35, *Median* = 3, SD = 1.83). Figure 4.6 shows that responses were relatively distributed and involved the full range. Although the most common response was 1 (*nobody*; 22%), this is closely followed by 3 (*some people*; 20%). Overall, the majority of respondents (78%) have encountered others discussing the concept metalinguistically, hinting at the prevalence of this ideology in some form.

Despite high average familiarity with the concept, participants tended to express a modest success rate in terms of their ability to accurately identify Asian Americans based on speech alone (M = 3.43, *Median* = 3, SD = 1.28). At the same time, all but one respondent (98%) felt that they could at least occasionally identify an Asian American speaker as such. As seen in Figure 4.6, the vast majority of responses clustered in the middle range of the scale, selecting either 2 (*occasionally*; 25%), 3 (*sometimes*; 32%), 4 (*about half the time*; 18%), or 5 (*frequently*; 18%). This shows that, while not a perfect ability, respondents overwhelmingly believed that it was possible to reliably infer Asian American identity from linguistic cues.

These results broadly confirm that the idea of identifiable Asian American speech patterns exists at a conscious level for many Asian Americans from California. This concept is available to discuss at a holistic level, and one of the sources of this ideology may be metalinguistic commentary.

4.5.2 Linguistic Features

With the aim of digging deeper into the details of this ideology, one of the main questions this survey intended to explore was: "What linguistic features are consciously linked to an 'Asian American' category, and to what degree of awareness?". Table 4.5 lists each of the 14 linguistic feature descriptions presented to participants along with their mean FAMILIARITY and ACCURACY ratings; features are ordered by the average of both measures. Figure 4.7 and Figure 4.8 visualize the model estimates and pairwise comparisons for each feature ordered by the mean rating of each scale.

Overall, ratings were highly variable across participants. Responses for all features spanned the full range of 1–7. As Table 4.5 shows, the range of average ratings is also relatively narrow and centered on the middle of each scale, falling between 3–5. For FAMILIARITY, this corresponds to roughly moderate familiarity for all features, suggesting that all statements were familiar to some extent but not extremely so. For AWARENESS, this signals that respondents, as a group, neither agreed nor disagreed with each statement, or else only converged on slight disagreement or slight agreement. In terms of relative differences, the top ranking features on both scales were the same: TH, OU, and PITCH. The lowest ranking features differed between the two scales. Features with the lowest FAMILIARITY were FRY, STACCATO, PAUSE and BREATHY while features with the lowest ACCURACY were PAUSE, FRY, BREATHY, and NASAL.

Model results support many of these observations and provide some additional insight. First, estimated marginal means from the FAMILIARITY model (Table 4.6) confirm that ratings for all 14 features differed significantly from zero (representing *not familiar at all*). Second, the estimated marginal means from the AWARENESS model (Table 4.7) find that most features did not differ significantly from zero (representing *neither agree nor disagree*). The two features that did, namely OU and TH, elicited likely agreement. Together, this suggests that while there was evidence that all feature statements were considered to be familiar to some extent, there was only reliable evidence that /oʊ/ and /ð/-stopping were consistently believed to

Feature	Statement	Familiarity	Accuracy
TH	Their “th” sounds more like “d” (e.g. “those”)	4.37 (2.15)	4.82 (1.73)
OU	Their “o” sounds rounder or different (e.g. “so”)	3.98 (2.14)	4.58 (1.79)
Pitch	They speak with a higher pitched voice	3.97 (1.89)	4.35 (1.63)
Quiet	They speak quieter	3.95 (1.96)	4.27 (1.75)
Uptalk	They end statements in a question (i.e. a rising pitch)	3.77 (1.84)	4.30 (1.62)
TAsp(iration)	They articulate “t” sounds more strongly/clearly (e.g., “to”, “wanting”)	3.83 (2.28)	4.19 (1.95)
TDel(etion)	They drop “t”s at the end of words (e.g., “can’t”)	3.78 (2.23)	4.17 (1.92)
Slow	They speak slower	3.58 (2.04)	3.88 (1.72)
Nasal	They sound more nasal	3.37 (2.12)	4.00 (1.80)
Like	They say “like” a lot	3.30 (2.09)	3.83 (1.86)
Staccato	They sound more staccato or clipped	3.18 (2.01)	3.93 (1.79)
Breathy	They speak more airy or breathy	3.27 (1.93)	3.83 (1.65)
Pause	They pause more	3.27 (2.11)	3.65 (1.78)
Fry	They speak with more vocal fry	3.07 (1.89)	3.68 (1.73)

Table 4.5: List of features and the wording of statements presented to participants, along with its mean (*SD*) familiarity and accuracy rating.

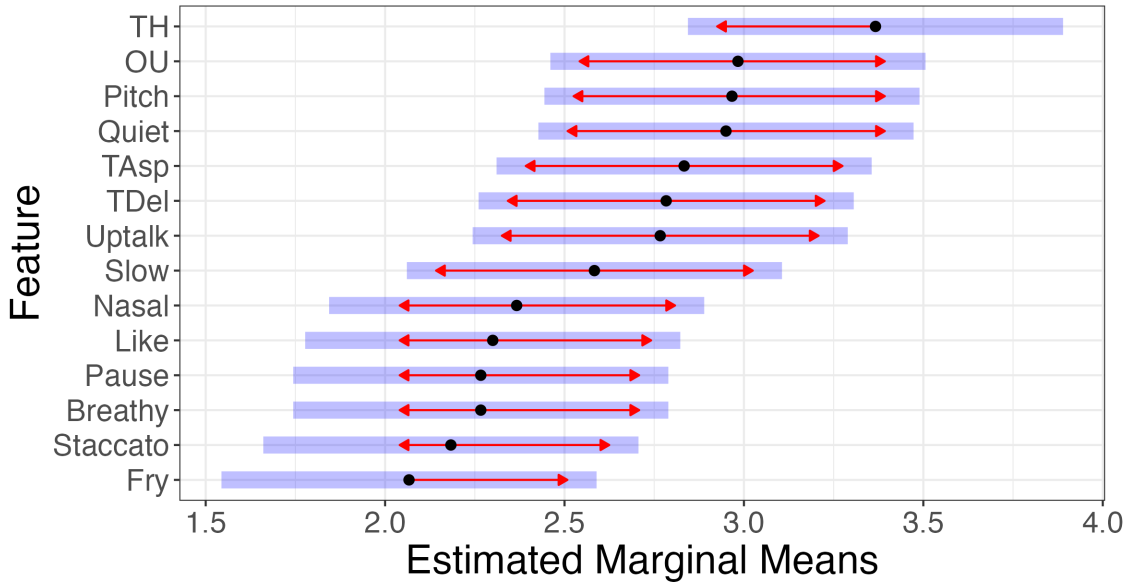

Figure 4.7: Estimated marginal means of FAMILIARITY ratings for each feature. Bars represent 95% confidence intervals and arrows represent approximate significance of pairwise comparisons based on Tukey-adjusted p values ($\alpha = .05$).

Feature	**Estimate**	***SE***	***df***	***t***	**CI lower**	**CI upper**	***p***
TH	3.367	0.265	188.570	12.705	2.844	3.889	$< .001$
OU	2.983	0.265	188.570	11.258	2.461	3.506	$< .001$
Pitch	2.967	0.265	188.570	11.195	2.444	3.489	$< .001$
Quiet	2.950	0.265	188.570	11.132	2.427	3.473	$< .001$
TAsp	2.833	0.265	188.570	10.692	2.311	3.356	$< .001$
TDel	2.783	0.265	188.570	10.503	2.261	3.306	$< .001$
Uptalk	2.767	0.265	188.570	10.440	2.244	3.289	$< .001$
Slow	2.583	0.265	188.570	9.749	2.061	3.106	$< .001$
Nasal	2.367	0.265	188.570	8.931	1.844	2.889	$< .001$
Like	2.300	0.265	188.570	8.679	1.777	2.823	$< .001$
Breathy	2.267	0.265	188.570	8.554	1.744	2.789	$< .001$
Pause	2.267	0.265	188.570	8.554	1.744	2.789	$< .001$
Staccato	2.183	0.265	188.570	8.239	1.661	2.706	$< .001$
Fry	2.067	0.265	188.570	7.799	1.544	2.589	$< .001$

Table 4.6: FAMILIARITY model estimated marginal means.

be accurate features of an “Asian American accent”.

Pairwise comparisons further reveal relative differences between features. TH was clearly the most familiar feature, being rated significantly higher on FAMIL-

IARITY than six of the lowest-rated features on this list: FRY ($p < .001$), STACCATO (p = .001), BREATHY (p = .002), PAUSE (p = .002), LIKE (p = .004), and NASAL (p = .011). In addition, three features—OU (p = .033), PITCH (p = .041), and QUIET (p = .050)—were rated as significantly more familiar than FRY, the least familiar feature. This suggests that while /ð/-stopping appears to be a particularly familiar feature in terms of ideologies about "sounding Asian", /oʊ/(-backing), higher pitch, and quieter speech also seem to be relevant to these ideologies.

With regards to impressions of ACCURACY, both TH and OU were rating significantly higher than PAUSE ($p < .001$; p = .003), FRY ($p < .001$; p = .005), and LIKE (p = .001; p = .048). Moreover, TH alone was rated as significantly more accurate than four additional features: BREATHY (p = .001), SLOW (p = .003), STACCATO (p = .007), and NASAL (p = .020). In this case, both /ð/-stopping and /oʊ/(-backing) appear to be relatively prominent linguistic features tied to "Asian" speech ideologies while other potential features like vocal fry or breathiness were not as likely to be endorsed as an aspect of Asian American speech.

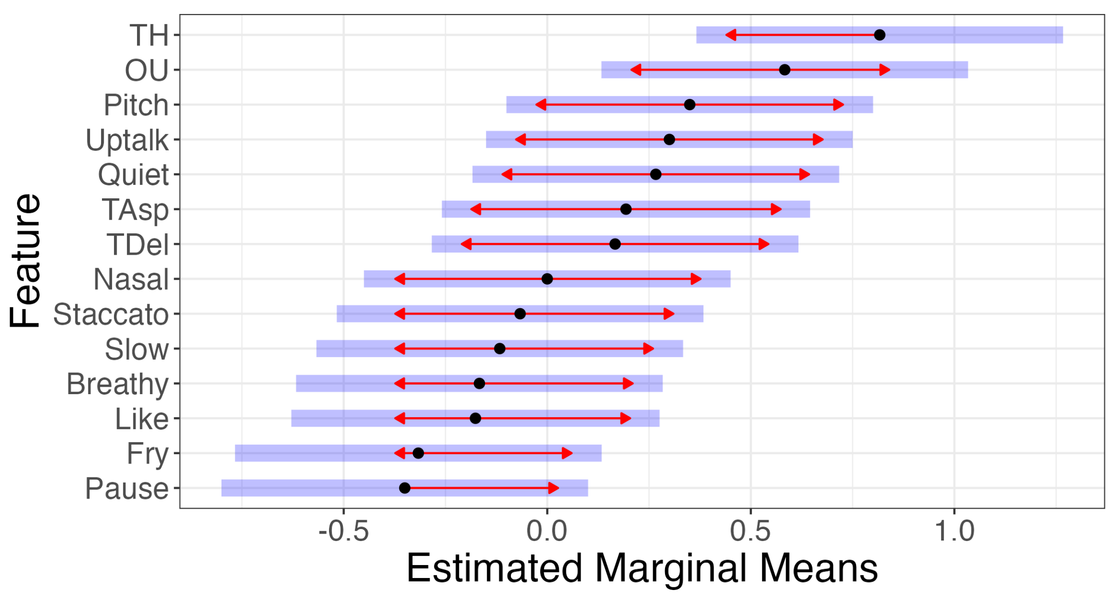

Figure 4.8: Estimated marginal means of ACCURACY ratings for each feature. Bars represent 95% confidence intervals and arrows represent approximate significance of pairwise comparisons based on Tukey-adjusted p values (α = .05).

Feature	**Estimate**	***SE***	***df***	***t***	**CI lower**	**CI upper**	***p***
TH	0.817	0.228	183.338	3.578	0.366	1.267	$< .001$
OU	0.583	0.228	183.338	2.556	0.133	1.034	.011
Pitch	0.350	0.228	183.338	1.534	-0.100	0.800	.127
Uptalk	0.300	0.228	183.338	1.314	-0.150	0.750	.190
Quiet	0.267	0.228	183.338	1.168	-0.184	0.717	.244
TAsp	0.193	0.229	186.179	0.844	-0.259	0.646	.400
TDel	0.167	0.228	183.338	0.730	-0.284	0.617	.466
Nasal	0.000	0.228	183.338	0.000	-0.450	0.450	1.000
Staccato	-0.067	0.228	183.338	-0.292	-0.517	0.384	.771
Slow	-0.117	0.228	183.338	-0.511	-0.567	0.334	.610
Breathy	-0.167	0.228	183.338	-0.730	-0.617	0.284	.466
Like	-0.176	0.229	186.179	-0.769	-0.629	0.276	.443
Fry	-0.317	0.228	183.338	-1.387	-0.767	0.134	.167
Pause	-0.350	0.228	183.338	-1.534	-0.800	0.100	.127

Table 4.7: ACCURACY model estimated marginal means.

Considering all the results at once, if degree of familiarity and agreement with feature-based descriptions of an "Asian American accent" can be interpreted as facets of awareness—that is, conscious access to sociophonetic associations—these results indicate that Californian Asian Americans indeed show some level of awareness for /oʊ/(-backing), following predictions. Further, /ð/-stopping emerged as an additional feature ideologically linked to "sounding Asian", with a similar or slightly higher level of awareness compared to /oʊ/(-backing). Numerically, /ð/-stopping was consistently considered to be more familiar and accurate, aligning with expectations that /oʊ/(-backing) would not be the most salient feature. In light of these results, I take a closer qualitative look at the data for these top two ideological features associated of Asian American speech to understand the fuller picture of awareness. To illustrate the distribution of ratings, Figure 4.9 displays the proportion of responses per Likert scale point for OU and TH.

I first start with the feature of interest, /oʊ/-backing. Figure 4.9 shows that the most common response for FAMILIARITY was 1 (*not familiar at all*; 22%), while other responses were relatively evenly distributed across the scale (78% were familiar to some extent). On the other hand, the most common response for ACCURACY was 6 (*agree*; 30%). Including those participants, over half of the sample (57%) agreed to some extent that /oʊ/ pronunciation was "rounder" or "different" in Asian American speech. At the community level, then, /oʊ/ is not necessarily readily recognized by all respondents as a feature of "sounding Asian", but individually, around half of participants considered it familiar and accurate.

Due to the lack of non-technical vocabulary to express /oʊ/-backing, it is unclear exactly what aspect of /oʊ/ vowels participants may have been envisioning in response to the statement that /oʊ/ is "rounder" or "different". Their impressions could be sourced from /oʊ/-backing, /oʊ/-monophthongization, the combination, or some other aspect. While open-ended responses across the survey do not answer this question, they do provide some more context as participants describe certain talkers' /oʊ/ as "round[ed] out", "elongated", "drawn out", "dragged

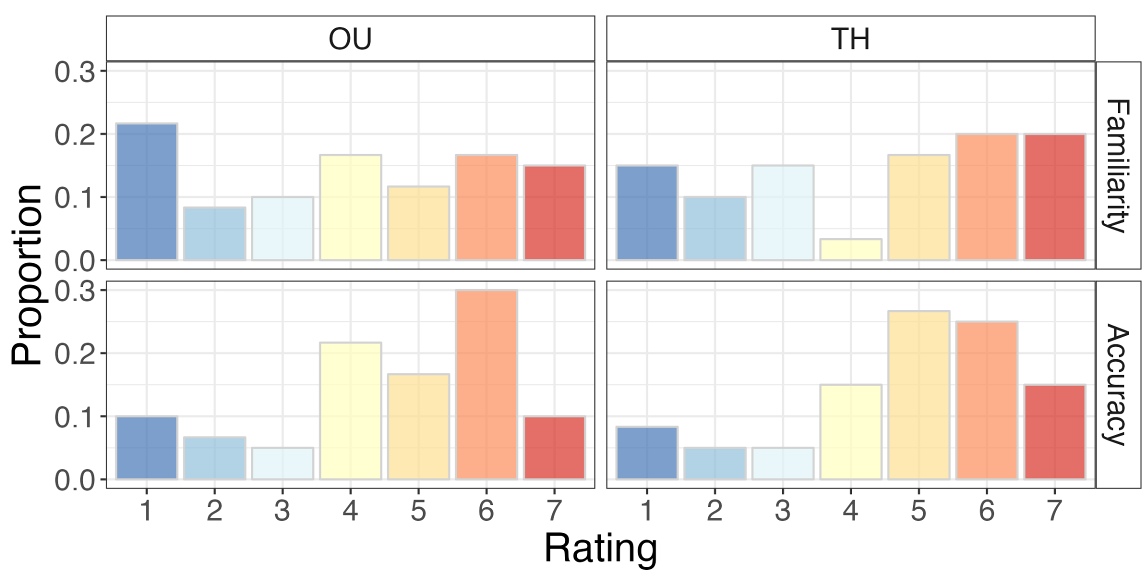

Figure 4.9: Proportion of participants who selected each point on 7-point Likert scales of FAMLIARITY and ACCURACY for /ou/ (OU) and /ð/ (TH) features.

out", and "lower in tone". This impression of a rounder, longer, and/or lower /oʊ/ could potentially be consistent with backing and monophthongization. As discussed in the rating task results, the presence of metalinguistic commentary further demonstrates that some participants were aware of /oʊ/ phonetic variation enough to discuss it explicitly, which supports the possibility that they used this feature to assess likelihood of a talker being "Asian".

Next, I consider /ð/-stopping, the top-rated feature. As seen in Figure 4.9, there was a somewhat bimodal distribution of FAMILIARITY ratings but the most common responses were 6 (*very familiar*; 20%) and 7 (*extremely familiar*; 20%). Overall, the greater part of respondents (85%) reported that /ð/-stopping was a familiar idea to some extent, and more than half (57%) felt that it was beyond moderately familiar (i.e.,ratings > 4). Moreover, two-thirds of participants (67%) agreed that /ð/-stopping was accurate to a certain degree. Evidently, a clear majority of respondents were familiar with the idea that /ð/-stopping is a part of "sounding Asian" and also believed it to be true.

These response patterns suggest somewhat wide-spread awareness for /ð/-stopping. It also hints at the potential for this feature to have influenced "Asian" ratings in the auditory rating task—in addition to, or even instead of, /oʊ/-backing. This possibility is supported by unprompted open-ended commentary pointing out /ð/-stopping as a noticeable feature of various talkers' speech during the auditory rating task, as well as by mentions in free responses to the post-task question about what participants "listened for" to guess ethnicity.

4.5.3 Social Types

The final question to be explored in these data are listeners' conception of "types" or subcategories of Asian Americans, especially whether "types" relating to a dimension of "mainstream" versus "ethnic" (linked to ideas of localness and foreignness) emerge from commentary and whether specific speech styles are associated. To do this, I qualitatively coded open-ended responses to a social category question and a speech style question. To remain focused, I highlight a selection of the most relevant results to understanding the predominant ideologies.

Of the 60 total social category responses, 48 participants offered at least one type of Asian American; most provided multiple examples, resulting in 147 total data points and 21 codes. Two of the most commonly-mentioned Asian American social types were the "FOB" Asian (n = 15) and the "Whitewashed" Asian (n = 12). The "FOB" Asian was broadly described as someone who has a closer affiliation with their ethnicity or Asian culture than the local American culture, linked to recent immigration and sometimes involved a negative valence. Appearance (e.g., fashion), actions (e.g., media consumption), and language styles (e.g., accentedness) were also invoked as part of the "FOB" character, exemplified by the following responses:

> "The next is the 'F.O.B', or Fresh-Off-The-Boat. These people tend to have a different *fashion sense that is distinctly not American* and will often *converse in an Asian language or carry an accent when speaking English*. They adhere to their Asian culture more than having adopted Western culture." (emphasis my own)

> "Asian Americans that hang on tightly to the foreign culture they originate from. I can't think of a politically correct term for this but these groups of people usually have *heavy accents, aren't fluent in English, consume Asian inspired media and adopt Asian inspired fashion*" (emphasis my own)

On the other hand, the "Whitewashed" Asian character was described briefly as those who "act white" or "American". Alternatively, this was implied using the common metaphorical terminology of “bananas” or “twinkies”, defined by multiple participants as “yellow on the outside but white on the inside”. This social type was also sometimes commented on with negative affect, such as being described as “too whitewashed”. Two other codes were used to characterize “More Asian” ($n = 4$) and “More American” ($n = 4$) types of Asian Americans. These were similar to the previous two but were placed in separate codes because they more broadly and neutrally indicated orientation towards “Asian” (e.g., “love their [ethnic] culture”) and “American” (e.g., “American Born Chinese”) identities and did not invoke “immigrant” or “white” associations. “More Asian” and “More American” could be viewed as umbrella categories that encompass “FOB” or “Whitewashed” Asians, respectively, or alternatively, as separate categories representing degree along a continuum of orientation (e.g., “FOB” > “More Asian” > “More American” > “Whitewashed”), in line with the “bicultural middle” discussed by Pyke and Dang (2003). Regardless, these codes can be interpreted as representing a dichotomous tension between “ethnic-oriented” and “mainstream-oriented” Asian American social types that arise from ideological commentary.

Responses to the speech style question provide some additional support for linguistic elements associated with these social dimensions. In these data, 29 participants provided relevant comments, resulting in 40 unique coded instances that involved 13 of the previously-identified social category codes. Some participants mentioned the “FOB” social type directly or indirectly ($n = 5$), and three of these mentioned the association with an “accent” or sounding “foreign . . . as in [E]nglish is not your first language”. This aligns with the unprompted speech style comments in some of the general descriptions for “FOB”.

Two participants also mention that they believe there is a specific style or “voice” associated with “regular ‘Asian American[s]’” or “second generation or later Asians”. Although no linguistic descriptors were provided, one participant commented that the speech style was "harder to discern", which hints at some level of general awareness without the ability to describe it. Although the texts don't provide much context, they could be interpreted as connecting to the "More Asian" and "More American" social types that represent some form of intermediate or less extreme orientation, paired with less distinctive or easily-described style. Specifically, the "regular" Asian American terminology was linked to an Asian American character who "fulfills all the typical Asian stereotypes" such as being smart, studious, and enjoying Asian cuisine, which I coded as part of a "More Asian" type, while "second generation" brings to mind the "American born Chinese" comments from the "More American" code.

Two comments mentioned the "Whitewashed" Asian type, and both identified these individuals as sounding in some way "white", but in different—and gendered—ways. Notably, one suggests that the speech style is a combination of "white bro and [A]sian [A]merican", while the other appears to group "Whitewashed" with "valley girl" social types, and specifies that the use of "vocal fry" by these types is linked to whiteness.

The latter is interesting as various other comments also brought up "Valley Girl" speech style but associated specifically with Asian Americans broadly (e.g., "I think any Asian American kind of sounds like a Valley Girl", "I think the 'valley girl' speech pattern is pretty common") or a type of Asian American (e.g., "I have noticed a valley girl Asian American accent", "Southern California Asian Americans... noticeable when I hear the 'Valley Girl' accent being used"). From these data, it is clear that some Californian Asian Americans have an association between what can be perceived as a "Valley Girl accent" and Asian American speech. What remains unclear is what linguistic features are relevant (aside from the named vocal fry), and how the Valley Girl persona or speech style is linked to racioethnic orientation within the context of Asian American speech. For example, is the use of Valley Girl elements always associated with whiteness (e.g., "White-washed" Asians), or is there a specific Asian American style that draws from Valley Girl elements but is associated with (local, Californian) Asianness (e.g., "More American" Asians, or even "More Asian" Asians)?

4.6 Discussion

This study employed an exploratory survey with a social evaluation task to assess sociolinguistic ideologies about AACs and their speech, framed as an "Asian American accent" or "sounding Asian (American)". A key focus was placed on investigating the role of /oʊ/-backing as an ideological cue to identifying Asian American speakers, along with the degree of awareness attached to it. Secondary aims were to identify other potentially relevant linguistic cues at varying levels of awareness, as well as to probe the links between intraethnic Asian American personae and "sounding Asian". In the following, I discuss first the findings related to awareness of linguistic cues and AAC ethnic identification, then touch on the connections to AAC personae varying in ethnic orientation.

4.6.1 Ethnic Identification & Awareness

The perceptual social evaluation task found that Asian American Californians were able to guess the ethnic identity of Asian Americans talkers from listening to short excerpts of speech. In this study, AAC talkers were on average rated as likely to be Asian, more so than comparison talkers, corroborating previous research (e.g., Hanna, 1997; Newman & Wu, 2011; A. Cheng & Cho, 2021). Plus, similar to previous studies (e.g., Newman & Wu, 2011), listeners were not necessarily accurate in identifying the specific ethnicity of speakers. This result substantiates the conclusion that listeners have linguistic ideologies tied to a pan-ethnic "Asian" or "Asian American" identity, even if there may be no pan-ethnic variety of Asian American English in reality.

As expected, some individual speakers were rated as more likely to be Asian than others. To contribute to answering the question of why, results show that speakers who produced more backed /oʊ/ vowels were perceived as more likely to be Asian, and less likely to be white, Black, or Latin/Hispanic. Due to the naturalistic source of the auditory stimuli, it is reasonable to expect that other features are co-occurring with /oʊ/-backing in the speech samples, and either those features, or the combination of/oʊ/ with those features together, drive perceptual identification responses(see discussion in Newman & Wu, 2011). Thus, while these results do not demonstrate whether or not /oʊ/-backing plays a specific role as a perceptual cue for Asian American ethnic identification, it is fully consistent with this possibility (see also P. Wong & Babel, 2017). At the very least, it shows that /oʊ/-backing appears to be linked to other features of "sounding Asian", and is a plausible candidate to be a socially-indexed marker of AAC identity for both speakers and listeners, adding to supporting evidence from production studies (e.g., Bauman, 2016; L. Jeon, 2017; A. Cheng, 2020).

Results from questions explicitly asking about ideology provide additional support for these findings. Generally speaking, most of the Asian American Californians surveyed in this study believed that Asian Americans can have a certain speech style different from other Americans, but report low to moderate consistency in actually identifying voices as Asian American. This lines up well with the variable by-talker perceptual identification behavior described above.

In terms of the awareness of specific linguistic features, /ð/-stopping and /oʊ/ (-backing) emerged in this study as the two linguistic features considered both most familiar and most accurate with regards to their association with "sounding Asian". In the listening task, these two features were also brought up multiple times as aspects of the speech that stood out to listeners, especially notable in responses to the two speakers perceived to be most likely Asian. Results overall suggest that respondents may be slightly more aware of /ð/-stopping than /oʊ/-backing, but differences were not particularly large. I interpret this as evidence for relatively wide-spread community-level awareness for these two phonetic variants as sociophonetic variables tied to an "Asian" or "Asian American" identity, at least within the Californian Asian American community. Of importance is the nuance that both features involved high variability at the individual level such that even when awareness was present, it ranged in degree from low to high. Altogether, these findings contribute converging evidence that /oʊ/-backing as well as /ð/-stopping are perceptual cues leveraged by some listeners to identify voices as "Asian" or as a specific Asian ethnicity, greatly expanding our understanding of the role of these variables in ethnic identification (P. Wong & Babel, 2017; A. Cheng & Cho, 2021) and ideologies about Asian speech (Chun, 2004).

To add, /ð/-stopping is an interesting case as it is a feature noted as commonly sourced from substrates in English. It appears as an ethnically-linked feature in many American communities, including African American and Chicano Englishes (e.g., Thomas, 2007; Eckert, 2008b), as well as in Asian-associated contact varieties, including Hong Kong English and Korean-accented English (e.g., Sewell & Chan, 2010; Schirra, 2012). In Chun's (2004) description of Mock Asian, /ð/-stopping was notably used in comedic performances to portray not only Chinese, Japanese, and Korean ethnicities but also unspecified Asian ethnicities. This suggests that /ð/-stopping could be a particularly marked ethnolinguistic cue in a North American English-speaking context, including Asian identity as one possibility in the indexical field. It is worth mentioning that /ð/-stopping has thus far not been specifically targeted as a feature in perceptual identification studies of AAC identity, nor in studies of AAC speech production. The current results provides strong motivation for future research to investigate the ideological, perceptual, and production-based links between /ð/-stopping and "sounding Asian".

Some other feature-specific results appear notable in light of prior findings and would be interesting to explore more systematically in future research. For example, the idea of Asian Americans speaking with higher pitch was ranked third in both familiarity and accuracy impressions, just below /oʊ/ and /ð/-stopping. This feature was also mentioned in free responses to the post-task question about strategies for guessing ethnicity. Since higher pitch has been previously reported to influence perceptual identification for Asian American women in particular (A. Cheng & Cho, 2021), this might suggest that it was also an ideological cue that influenced listeners' perception of Asian speaker identity in this study, which included only voices perceived as women. Whether or not these beliefs are based on actual evidence or experiences with Asian American speech production remains to be seen.

On the contrary, speech rhythm, at least relating to the description of sounding "staccato" or "clipped", did not appear to be particularly familiar or agreed with relative to the other features. Its placement among other potential ideological cues to AAC identity within the current study is somewhat surprising given that speech rhythm has been directly examined in both production and perception studies of AAC speech (e.g., Newman & Wu, 2011; Bauman, 2016). The specific descriptions used in the current study were taken from previous metacommentary results (Hanna, 1997; A. Cheng & Cho, 2021), meaning these elements have been found to be consciously associated with Asian American speech, and to such an extent that it was offered up in open-ended discussion without prompting. However, it is possible that the interpretation of the descriptors by individuals participants (and the researcher) vary such that, for example, "clipped" indicates different features to different people and does not refer to the same type of impression as "staccato".

The resulting perceptual impressions further may or may not in fact link to speech rhythm as conceptualized by researchers. Relatedly, breathiness was identified in this study as a relatively unfamiliar and inaccurate description of AAC speech in terms of ideologies, but was identified as a potential cue to AAC identity in Newman and Wu (2011). Additional study will be needed to understand whether these disconnects are due to imprecise assessments and/or imperfect links between descriptors and phonetic detail, differences in speaker or listener populations, or certain features being markers of AAC identity outside of awareness.

At the same time, the fact that all statements were rated moderately familiar on average—with individuals using both ends of the scale—suggests that there may simply be a large number relevant features to AAC identity, both ideological and perceptually, of which only a subset is relevant to each individual. This would align with the notion that all listeners have their own ideological constellation of features associated with "sounding Asian", and thus different people may be aware of and attend to different features (Newman & Wu, 2011). The source of these differences could be related to listeners' individual experience with particular AAC ethnicities that are linked to certain features over others, and this type of experience may vary systematically by listener background, including ethnicity and location. The AAC-linked features that were identified with the most consensus—/oʊ/-backing and /ð/-stopping—may have been ones that crossed ethnic boundaries, either being associated with multiple specific ethnic groups and/or associated with "Asian American" as a construct (e.g., Chun, 2004). Since all feature statements included in the current study were selected with the potential to be linked to an "Asian American accent", there were no statements to act as a baseline representing clear lack of sociolinguistic association (e.g., cues that are clearly associated with other varieties). Future work examining specific linguistic cues to AAC identity may wish to more thoroughly consider the role these factors play in shaping ideologies about "Asian American" speech.

4.6.2 Ethnic Orientation & Personae

In the listening task, the presence of /oʊ/-backing appeared to be evaluatively associated with accentedness and foreignness, specifically predicting increased perception of talkers being accented and having grown up outside of the U.S. or Canada. This shared result points to an ideological link between Asian identity, being foreign, and having a foreign accent (Tuan, 1998; Reyes & Lo, 2009), held by Asian Americans as well as others. It also suggests the potential for /oʊ/-backing—as well as the co-occuring /ð/-stopping—to be specifically associated with "foreign" Asians, in line with descriptions of stereotyped varieties of English (Chun, 2004).

Foreignness was also a key element of the ideological space in terms of intraethnic subcategories of Asian Americans. Participants expressed many ideas about different social types of Asian Americans that intersect with linguistic styles, including particularly clear ideologies about "FOB", culturally Asian, or ethnic-oriented Asian Americans in opposition to "Whitewashed", culturally American, or mainstream-oriented Asian Americans. The "FOB" persona described by this diverse set of Asian American Californians is linked to not only certain visual styles like fashion but also linguistic styles tied most closely to foreign-accentedness and non-native English. The "Whitewashed" persona does not draw as much commentary on linguistic styles, but is described to elicit an impression of "sounding white", seemingly linked to stereotypical Valley Girl speech styles (see Bucholtz et al., 2007, on the ideological link between Valley Girl, whiteness, and Southern California).

These ideological constructs line up surprisingly well with prior research conducted over two decades ago in the 1990s (Pyke & Dang, 2003), demonstrating the prevalence of ethnic-oriented and mainstream-oriented Asian American personae available to Californian Asian Americans up until the present day. These findings support the position that a salient dimension within broadly circulating sociolinguistic ideologies about Asian Americans in California is represented by a continuum between orientation towards "ethnic", "foreign", or "Asian" culture on one end (tied to "foreign accents"), and "mainstream", "American", or "white" culture on the other end (tied to "sounding white").

The perceptual evaluation task found more extreme /oʊ/-backing to result in stronger and more consistent impressions of the speaker being Asian, foreign, and accented. While this may be due to other a variety of factors, such as co-occurrence with a greater number of other Asian-associated features, it could potentially suggest that perceptual cues to narrower Asian identities fall along this continuum as well. That is, particularly backed /oʊ/ might elicit impressions of a "FOB" or ethnic-oriented persona while a moderately backed /oʊ/ might elicit a more "bicultural middle" speaker impression. If so, this would add more complexity to the task of determining what cues lead someone to "sound Asian", as we would further need to consider both the strength of the linguistic cue and the type of Asian that is invoked via ideological associations. Alternatively or in addition, the ideological strength of each cue tied to a particular persona may need to be accounted for, such that /oʊ/-backing or /ð/-stopping could be linked strongly to a foreign-associated macrosocial "Asian" or "FOB" category, but linked weakly or not at all to more mainstream-oriented personae like the "Normal Asian". The current study does not tackle this question; more research will be needed to determine the full picture of AAC social and linguistic ideological associations.

Before moving on, I pause here to consider the ideological interaction of the Valley Girl persona with Asian American identities. This connection was raised not only in the current study but also in metalinguistic commentary from previous research (A. Cheng & Cho, 2021). Californian Asian American participants in this study often linked Valley Girl to an "Asian American accent", but one participant also linked Valley Girl features to a specific "Whitewashed Asian" persona, including mention of vocal fry as a part of this style. A. Cheng and Cho's (2021) Korean American respondents, both from California and not, linked Asian Americans to Valley Girl with reference to high rising terminal pitch contours and use of the filler word "like". However, in the perceptual rating task, especially in the metalinguistic commentary of particular speech samples, these Valley Girl-related features seemed most strongly associated with speakers who were less identified as Asian. Moreover, two of these features were rated relatively low on familiarity and accuracy in regards to AAC speech—especially vocal fry, but also use of "like". High rising terminal, or uptalk, was the only one that might have been more closely associated in awareness with Asian American identity (also mentioned in metalinguistic commentary in Hanna, 1997). For now, the exact ideological link between Valley Girl features and Asian American speech remains unclear. This intersection seems to be a relevant piece of the puzzle to understand what "sounding Asian (American)" means to listeners, and may be fruitful for future research to investigate.

4.6.3 Conclusion

In all, these results provide a clear opportunity to examine the role of awareness and personae in sociophonetic perception and representations, using the context of Californian Asian American speech. The phonetic variables of /oʊ/-backing and /ð/-stopping may be linked—with stronger or weaker awareness for different listeners—to perceptual impressions of Asian (or AAC) identity, accentedness, and foreignness. They may also be linked specifically to a recognized "FOB" or ethnically-oriented AAC persona, instead of to a general "Asian" macrosocial category. Before moving on to investigations of ideology's influence on speech processing, the ideas that listeners have about relevant social types of AACs and how they speak must be more clearly defined. This is because personae may be important social constructs at the level of individual cognition, underlying patterns of behavior attributed to macrosocial category constructions (Eckert, 2008a; D'Onofrio, 2020). These ideas are tied not only to linguistic styles but also to visual details, which interact with each other during speech perception. As such, the next chapter focuses on operationalizing these microsocial AAC categories from multiple angles, including examination of both visual and auditory associations.

CHAPTER V

Looking and Sounding Asian: AAC Ideologies in Social Evaluation

5.1 Introduction

To follow up on the qualitative results from Chapter IV which also involved a broader scope of inquiry, this chapter reports on a series of three targeted surveys that together build a more detailed picture of the sociolinguistic ideologies surrounding AACs. I employ social evaluation tasks in both visual and auditory domains to describe the social impressions of AACs that vary in perceived acculturation. These findings provide context for the range of associations linked to AAC personae—including how they look and how they sound—as well as the extent to which AAC voices may be recognized by others without the inclusion of two potentially relevant phonetic cues to AAC identity, namely /oʊ/-backing and /ð/-stopping. In doing so, the results additionally serve to norm the visual and auditory stimuli that will be used in subsequent perception experiments.

5.2 Background

Despite a common focus on macrosocial categories in sociophonetic perceptual research, the nature of social information linked to phonetic detail in a perceiver's cognitive representations is much richer. This includes constructs like social personae that are tied to stylistic elements or clusters of features (e.g., linguistic, visual, behavioral) and can mediate individual experiences (see review in Section 2.1.3.1). In one example, MacFarlane and Stuart-Smith (2012) showed that several phonetic variants were linked to a "Glaswegian" persona (compared to a "Glaswegian Uni" persona) constructed based on local brand logos that indicated particular lifestyles practices, including what they wear and what they do (e.g., bars, clothing brands, sport leagues, and lunch spots). Hearing /l/ vocalization, for instance, strongly influenced listeners to identify the speaker as someone represented by the "Glaswegian"-related brands. In line with general trends, speech research on Asians in North America often focuses on the macrosocial construct of "Asian" and its ideological links to foreignness (see full review in Chapter 2.2). This alone, however, does not represent the full picture of how AACs are viewed.

As reviewed in detail in Section 4.2.2, there exist ideologies circulating among Asian Americans about different types of AACs that are variably aligned with their ethnic culture or American culture. Based on these shared ideologies, AAC community members may be categorized based on perceived degree of acculturation, separating more culturally Asian "FOB" Asians from bicultural "normal" Asians and more culturally American "whitewashed" Asians (Pyke & Dang, 2003). Each of these types have their own set of social associations (e.g., behavior or appearance) and may elicit different attitudes. In this way, we can see the "FOB Asian", "Normal Asian", and "Whitewashed Asian" as characterological figures that are tied to certain ways of being and speaking—in other words, personae (e.g., Eckert, 2008a; D'Onofrio, 2020).

The results of Chapter IV confirmed that many Californian Asian Americans in the present day have similar ideas about different kinds of AACs and how they speak, evidenced by their descriptions of AACs that I classified as "FOB", "more

Asian", "more American", and "Whitewashed" Asian types. While participants' comments from Chapter IV highlighted foreign-accentedness and non-American fashion as observable features of the FOB Asian persona (in line with previous explorations of the FOB construct; e.g., Pyke & Dang, 2003; M. Jeon, 2007), Normal Asian and Whitewashed Asian personae were not described with similar observable features. Though various appearance-related details, for example American-style or casual clothing, were contrasted against FOB styles by informants in Pyke and Dang (2003), these may or may not extend to the current population of Asian Americans.

In terms of ideologies about AAC speech, evidence for linguistic features that may be used by listeners to perceptually identify AACs is still inconclusive (see review in Section 4.2.1). Regardless, fully answering the question of what makes someone "sound Asian" will require more than simply identifying features associated with the generic macrosocial category. As discussed in Section 4.2.3, a major part of the unknowns about "sounding Asian" is the linkage between particular AAC voices or speech styles and more narrowly-defined AAC social types. Along these lines, some studies have focused on perceptual ethnic identification of specific ethnic groups, like Korean Americans (A. Cheng & Cho, 2021) or Chinese Canadians (P. Wong & Babel, 2017), and some reports hint at intersectional gender-based differences (Newman & Wu, 2011; A. Cheng & Cho, 2021).

Research on this topic has generally not examined ideologies of AAC speech in the context of differently-acculturated subcategories of AACs, though one study is directly relevant. In the process of norming perceptual stimuli, D'Onofrio (2019) conducted social evaluations on a similar culture-related distinction in two photos of the same East Asian man with different visual styles. In this case, one photo was evaluated as relatively less likely to be from the US, less American, more foreign-accented, and more nerdy. This would align with characteristics of the FOB Asian persona, especially in terms of the accent associations, while the other photo was perceived as more American. Still, the reported outcomes may be limited to describing those specific photos and not general ideologies about AAC acculturation. The connections to speech so far also only touch on high-level ideas about accentedness. Generally speaking, beyond whether people can be identified by voice as ethnically Asian, the question of whether people can sound specifically like a FOB Asian, Normal Asian, or Whitewashed Asian persona—and what might lead to those impressions—is unexplored.

5.2.1 The Current Study

The question guiding this study is as follows: What are the detailed sociolinguistic associations that make up listeners' mental representations of differently-acculturated AAC social types, including visual styles, speech styles, and other social characteristics? Documenting the complex ideological landscape of AAC speech is important to making informed interpretations of how that speech is perceived by others. More specifically, the current study systematically examines whether different individuals, represented by photos and voices, can be interpreted as embodying Asian American personae with different cultural orientations, as well as how these personae may diverge in evaluation of various social characteristics.

Building off of the acculturation continuum described by Pyke and Dang (2003), I conceptualize perceived AAC acculturation as a continuum where "ethnic" and "mainstream" orientations are located on opposing ends. On one side, individuals are perceived as more ethnically-oriented or culturally Asian (e.g., the FOB Asian persona), and on the other, they are perceived as more mainstream-oriented or culturally American (e.g., the Whitewashed Asian persona). Individuals in the center are perceived as oriented to both the ethnic and the mainstream—that is, bicultural (e.g., the Normal Asian persona).

While broadly descriptive, this simple tripartite distinction may conflate multiple recognizable social types that differ gradiently in degree of cultural orientation within each region of the continuum. For example, non-American Asians or new immigrant Asian Americans might be grouped on the ethnic side of the continuum along with Asian Americans who grew up in North America but who are culturally-aligned with these groups. These sub-categories could possibly all be considered ethnically-oriented Asians (though only the Asian Americans are likely to be labeled as "FOBs"), but at what points on the theoretical continuum would these groups land? Not only that, where would we place the boundary between a distinctly ethnically-oriented Asian and a bicultural Asian who leans somewhat ethnically-oriented? The same issues arise on the mainstream side of the continuum when considering relatively mainstream-oriented bicultural Asians and Asians who are monoculturally American (and likely considered Whitewashed Asians). Nevertheless, despite lacking clarity in the details, the overarching structure of such a continuum seems to be ideologically relevant and offers a valuable starting point to investigate these constructs further.

For the remainder of this dissertation, I approach the interpretation of differently-acculturated AAC personae with relative flexibility. I mainly focus on the higher-level distinction between relatively "ethnic" Asians and relatively "mainstream" Asians, examining differences between types of people who are broadly perceived as contrasting on this continuum. As well, I refrain from making strong claims about the extent of difference between emergent types or exactly where on the continuum they would fall. To be able to concretize these ethnic-oriented and mainstream-oriented ideological constructs, I aim to link them to specific images and voices of AAC(-perceived) individuals using a quantitative approach.

The current study uses three surveys to identify and describe the social impressions of ethnically- and mainstream-oriented AAC types, including impressions of particular ways of speaking, via photos and voices. As in the previous chapter, this study samples Asian Americans from California, who are more likely to be familiar with the relevant ideological constructs. In the first survey, I investigate at a coarse-grained level whether different AAC-perceived photos can be associated with certain key social characteristics relating to the dimension of acculturation. In the second survey, I assess in more detail how the visual features of selected photos representing different AAC personae may elicit impressions of various social attributes and speech styles. In the third survey, I add in auditory speech samples to similarly examine how selected AAC voices might align with impressions of different AAC personae.

5.3 Survey 1a: Photo Selection

The goal of this first survey was to assess whether the visual styles portrayed by people in photos could broadly elicit impressions of different AAC social types along a continuum of acculturation. For stimuli norming purposes, this step served to narrow down the options for photos that would be able to represent the targeted AAC personae and comparison white personae.

In order to assess a large number of photos on a few key social dimensions, a photo selection task was used. Participants were given descriptions representing certain social characteristics of interest, including about race/ethnicity, accent, and cultural alignment. Then, they were shown a set of photos and asked to select all the people that they felt matched the description. By assessing how often each description was attributed to a particular photo, we can broadly describe the social impression of each photo, especially whether they were likely to be perceived as more mainstream-oriented or ethnically-oriented AACs.

5.3.1 Methods

5.3.1.1 Visual Stimuli

The visual stimuli consisted of 72 photos depicting a person's face.[1] Images were stock photos sourced from Pexels (https://www.pexels.com/), a stock photo website. All photos were freely available for use under a Creative Commons Zero license. Two main groups of photos were targeted: those depicting people who may be perceived as an "Asian American Woman" (AAW) and a matched set depicting people who may be perceived as a "White American Woman" (WAW).

In the first stage of photo selection, potential AAW photos were identified for consideration by searching for keywords like "Asian American", "woman", "face", and "portrait". Based on these initial seed photos, I additionally searched through related photos and other photos from the same photographer to find similar images in an iterative fashion.

Photos were then judged on a set of inclusion criteria. In general, images selected for inclusion were color photos clearly showing a face that could potentially be interpreted as a personal photo or professional portrait. Based on the pool of available options, all photos depicted smiling faces looking at the camera, as well as represented individuals who looked to be in their twenties or thirties. To assess inclusion of photos in the AAW group, I examined various elements of the image and photo subject, including clothing, hairstyle, makeup, facial expression, and background or location clues that aligned with more American- or Asian-associated cultural norms or trends. Taking these factors into account, I used my own intuitions as an AAC individual to provide a holistic judgment of whether the photo subject could potentially be interpreted as an East or Southeast Asian American woman, either with mainstream or ethnic orientation, to other AACs.

Photos that were judged to meet the criteria were downloaded and edited in

[1] Stimuli are available from the author on request.

Adobe Lightroom. All were cropped to a 4:3 ratio centering on the face and including some portion of the upper body. If needed, photos were further adjusted to achieve a similar level of brightness and to reduce the impression of stylized photography. The cropped photos underwent another round of judging against the inclusion criteria, which produced 36 photos of unique individuals tentatively intepretable as AAW. A second search process was then conducted for potential WAW photos. This followed a similar search method, with the addition of searching through related photos based on the selected AAW images. In the end, 36 photos tentatively interpretable as WAW were also selected.

5.3.1.2 Survey & Procedure

The survey was hosted on Qualtrics. Before taking part, participants were told that the study was about social perceptions of people based on style and that they would be asked to "look at photos of people and select the ones you think match a particular description". The main portion of the survey was the photo selection task. Participants were given the following instructions:

> In this task, you will see photos of different people. You will read an adjective or phrase, and you should click on every person that you think the description applies to. You can select as many or as few as necessary (e.g. all or none of the photos is possible).

There were four categories of descriptors presented in this order: Region, Race/ ethnicity, Accent, and Culture. The order was fixed so that participants could respond to the more general or concrete demographic descriptors first, followed by the more specific or abstract cultural ones.

Given that the participant sample was Californian and the stimuli were intended to represent peers, the Region category was included to screen for photos that could plausibly be interpreted as local. For this, all participants were asked

	Asian/Ethnic	White/Mainstream
Region	From California	
Race/ethnicity	Asian American	White American
Accent	Foreign-accented	American-accented
Culture	Participate in ethnic culture (e.g., traditions, food, media)	Participate in American culture (e.g., traditions, food, media)

Table 5.1: Descriptor labels by category (Region, Race/ethnicity, Accent, Culture), including three pairs of descriptors designed to assess Asian or ethnic-oriented associations versus white or mainstream-oriented associations.

to select photos in which the subject looked like they were "from California". The other three categories were designed to probe three elements relating the targeted AAC social types. In order to assess these in more detail, two opposing descriptors were included per category, one representing "white" or "American" impressions and one representing "Asian" or "ethnic" impressions (Table 5.1). For each category, half of the participants were randomly shown one label while the other half was shown the opposing label. This meant that participants would only see one label per category (e.g., "Asian American" or "White American") but could be presented with any combination of descriptors across the three categories (e.g.,"Asian American", "American-accented", and "Participate in ethnic culture").

The photo selection task was self-paced. Each participant saw four blocks, one per descriptor category. At the beginning of each block, participants were informed about the descriptor they would be using to judge photos. Then, they were presented with nine consecutive trial pages containing the descriptor along with eight photos. At the top of the page, participants were prompted with "Select all the people who look like they (are)", followed by the descriptor phrase in upper case letters (e.g., "FROM CALIFORNIA"). Below that were the eight photos arranged in two rows of four. All 72 photos were presented per block, meaning that each participant judged every photo on the four descriptors they saw.

Across the task, photos were presented in a pseudo-randomized manner. Ten randomly ordered lists of the 72 photos were generated ahead of time; these were then divided into nine subset lists of eight photos each. Before each block was presented to a participant, one of the ten lists was randomly selected. Within each list, the nine subset lists were randomized. Additionally, within each subset list, the eight photos were randomized. As a result, although the combinations of photos per page were restricted to ten possibilities, both photos and pages within a block were randomly ordered.

After completing the four blocks, participants were asked what they thought the task was about, what they thought about the task as they were completing it, whether they recognized any of the people, and whether they noticed anything else about the stimuli or task. Then, they completed a few questions about their demographic background, including age, gender, and ethnicity. The survey took on average 13 minutes to complete and each participant who did so received $2.80 in compensation.

5.3.1.3 Participants

In total, 40 Asian Americans from California completed the study remotely via Prolific (www.prolific.com). Participants were invited to participate on Prolific based on the following screening criteria: American, located in the US, identified as East Asian or Southeast Asian, and listed California as their state of birth and current state of residence.[2] The mean age was 30 (M=30.7, SD=9.8, range=20-63), and the sample included 21 women and 18 men (1 chose not to respond). All were living in California at the time, and all but one were born there (1 was born in Asia). In addition, all participants self-identified as Asian (22 East Asian, 16 Southeast

[2]Screener responses are sometimes inaccurate due to respondent errors or possible missing responses. Description of the sample is thus mainly based on demographic information provided in the survey itself.

Ethnicity	Total (n=40)	Man (n=18)	Woman (n=21)	NA (n=1)
Chinese	**8**	3	4	1
Filipino	**7**	4	3	0
Vietnamese	**6**	4	2	0
Cantonese	**4**	1	3	0
Japanese	**3**	1	2	0
Korean	**3**	1	2	0
Chinese-Filipino	**2**	0	2	0
Taiwanese	**2**	2	0	0
Chinese-Vietnamese	**1**	0	1	0
Filipino-Japanese	**1**	0	1	0
Filipino-Norwegian	**1**	0	1	0
Laotion	**1**	1	0	0
Thai	**1**	1	0	0

Table 5.2: Number of participants by self-identified ethnicity and gender.

Asian, 2 East and Southeast Asian), including two who identified as Mixed (see Table 5.2 for specific ethnicities). The most common ethnic backgrounds represented in this sample were Chinese, Filipino, and Vietnamese.

5.3.1.4 Data Analysis

Each of the 72 photos (36 AAW, 36 WAW) received 20 judgments per descriptor (i.e., one each from half of the participants), except for the Region descriptor which received 40 judgments per photo (i.e., one each from all participants). In total, 11,500 binary selection responses were analyzed.

To probe whether particular photos were more often selected in response to particular descriptors, I first calculated the proportion of selections per image and descriptor. That is, the total number of times a photo was selected by different participants in a descriptor block was summed then divided by the number of times it was presented with that descriptor (i.e., 20 or 40). This resulted in seven proportion values per photo for: "From California" (FrCA), "Asian American" (AsAm), "White American" (WhAm), "Foreign-accented" (ForAcc), "American-accented" (AmAcc), "Participate in ethnic culture" (EthCul), and "Participate in American culture" (AmCul).

In order to understand overall response patterns per photo across the seven descriptors, these variables were submitted to a principal components analysis (PCA) in R using the *PCA* function from *FactoMineR* package (Lê et al., 2008) and Varimax-rotated using the *Varimax* function from the *GPArotation* package (Bernaards & Jennrich, 2005). PCA allows us to both summarize the data to understand whether and how these social impressions can be grouped together, as well as functioning to reduce the data into a smaller number composite variables that can be used to describe the social impressions. Thus, this method provided two benefits: insight into ideologically relevant social dimensions for describing AAC social types and scores to categorize individual photos based on the relevant social dimensions.

To select the number of components, a number of diagnostics were used. Parallel analysis and the latent root criterion (i.e., factors with eigenvalues > 1) suggested retention of two components while the scree test (i.e., location of the elbow) and differences in eigenvalues between components (i.e., differences $>$ 1) suggested three components. Based on greater interpretability of the factors, three components were ultimately retained. Those three components accounted for 93.3% of the total variance. All variables had high communality values ($>$0.9), meaning all the variances per variable were well-accounted for by the factor solution.

Varimax rotation was applied to factor loadings to aid in interpretation. Loadings over 0.5 were considered in interpretation, following guidelines for selection of loadings that are considered to be practically significant (Hair et al., 2019). In the end, all variables had at least one significant loading, all variables' communality were high, and no variable has cross-loadings.[3]

5.3.2 Results

5.3.2.1 By-Group Selections

First, I describe the overall response patterns by reporting proportion of selections at the level of the assigned photo group. The distribution of proportions per photo for each of the seven descriptors are visualized in Figure 5.1, separated by whether photos were expected to represent AAW or WAW categories. If all participants selected a particular photo for a particular descriptor, the proportion would be one. Conversely, the value would be zero if no participants ever selected that photo and 0.5 if half of the participants selected that photo.

In general, demographic impressions of the photos were relatively balanced across the sample. As seen in Figure 5.1, the proportion of photo selections for FrCA was centered around the halfway point (M=0.53, SD=0.13) for both AAW photos (M=0.54, SD=0.13) and WAW photos (M=0.52, SD=0.14); photos in the two groups also covered a similar range of proportions. This indicates that all photo subjects were interpreted as from California at least some of the time, though no photos were ever unanimously considered to be Californian or not Californian. In addition, about half of the photos seemed more plausibly associated with Californian regional identity than others (M>0.5=0.56).

For the Race/ethnicity labels, responses patterned as expected: AAW photos were likely to be selected for AsAm (M=0.76, SD=0.12) and not likely to be selected for WhAm (M=0.04, SD=0.07) while the opposite pattern was true for WAW. This shows that many of the photos were interpreted as either Asian American or White

[3] If a minimal value of 0.4 was used to identify significant loadings, two variables would have had cross-loadings. To check whether this would be a problem, I assessed the ratio of variances following Hair et al. (2019). As the ratios were larger than 2.0, the weaker loadings were deemed to be ignorable, aligning with the final decision to exclude loadings below 0.5.

American, in line with the way the photos were originally grouped, though there was some degree of variation in agreement.

Where we see some more nuanced differences between photo groups is in the Accent and Culture categories. When judging ForAcc, different AAW photos were selected at different rates (M=0.37, SD=0.19) compared to WAW photos (M=0.14, SD=0.09), which tended not to be picked on the whole. Similarly, the AmAcc prompt elicited a much wider range of proportions for AAW photos (M=0.56, SD=0.22) than for WAW photos (M=0.78, SD=0.10), which in this case tended to be selected. These patterns suggest that the way certain AAW photos looked led Californian Asian Americans to think that the pictured individual was or was not likely to speak with an American accent (or a foreign accent). In other words, visual styles within a group of people interpreted as Asian American—based only on a photo—was associated with holistic expectations of speech style.

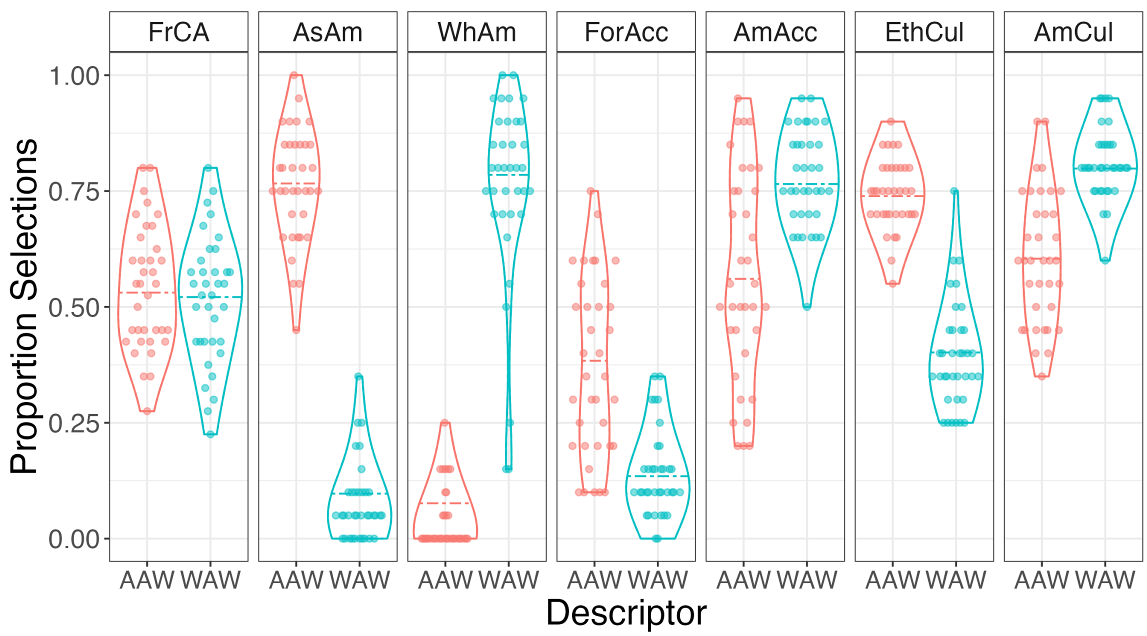

Figure 5.1: Violin plot showing the distribution of times a photo was selected per descriptor by photo group. Points represent proportions for each individual photo. Photos are grouped as Asian American Woman (AAW) or White American Woman (WAW) based on the author's judgment of their likely demographic interpretation.

In terms of Culture, AAW photos tended to be selected for EthCul (M=0.74, SD=0.07), unlike WAW photos (M=0.39, SD=0.12). On the other hand, AAW selection proportions for AmCul (M=0.61, SD=0.14) patterned similarly to the Accent descriptor conditions in terms of a greater range of variation compared to WAW photos (M=0.81, SD=0.08), which were also more concentrated in the higher proportions. Taken together, this seems to indicate that while all Asian American-perceived photo subjects were interpreted as people who would participate in ethnic culture, only some of these individuals were seen as also likely to participate in American culture at the same time.

5.3.2.2 By-Photo Component Scores

The factor loadings for the three components per variable from the Varimax-rotated PCA are reported in Table 5.3 while the individual factor scores (standardized) and loadings are visualized in Figure 5.2. Rotated components (RCs) were interpreted as follows.

The first rotated component (RC1) accounted for 40.40% of the variance in the

	Ethnic Background	Californian	American Upbringing
Variable	RC1	RC2	RC3
FrCA	-0.192	**0.934**	0.294
AsAm	**-0.943**	0.069	-0.238
WhAm	**0.920**	-0.129	0.327
ForAcc	-0.351	-0.238	**-0.852**
AmAcc	0.259	0.265	**0.878**
EthCul	**-0.839**	0.096	-0.452
AmCul	0.401	0.086	**0.854**
Eigenvalues	2.828	1.038	2.679
% Variance	40.401	14.824	38.271

Table 5.3: Varimax-rotated PCA factor loadings along with the sum of squares (eignvalue) and percentage of total variance explained per component (RC=rotated component). Each component is labeled with its interpretation. Component loadings > 0.65 are indicated in bold.

data. Significant loadings included WhAm in the positive direction and AsAm and EthCul in the negative direction. As such, RC1 was interpreted as a component linked to race/ethnicity or family background; I labeled it as the "Ethnic Background" dimension. Higher scores indicated that a photo elicited social impressions of being White American while *not* being Asian American or someone who participates in ethnic culture. As expected based on the construction of the stimuli, photos were differentiated on whether they looked Asian American or White American. However, this grouping of variables also suggests that ethnic cultural alignment may be linked to being perceived as Asian American, at least for social evaluative judgments on the current set of photos.

The third component (RC3) accounted for the second greatest amount of variance in the data (38.27%). This component was positively correlated with AmAcc and AmCul along with being negatively correlated with ForAcc. This led me to interpret RC3 as an American culture or upbringing-linked component, labeled as the "American Upbringing" dimension. Higher scores would indicate that a photo subject was perceived as someone who speaks with an American accent and participants in American culture while also not speaking with a foreign accent. This suggests that one important dimension that differentiated photos in this sample was whether someone was perceived as growing up in the US, including speaking American English and being culturally American. This pattern of results suggests that accent decisions were tied to location but not necessarily to ethnic activities.

The final component (PC2) accounted for 14.82% of the variance in the data and was defined by a strong loading of FrCA in the positive direction. This is straightforwardly interpreted as the "Californian" dimension, where higher scores are associated with a stronger impression of someone being from California.

The biplot in Figure 5.2 provides some additional insight by illustrating where individual photos fall along the two key dimensions. Visually, we can identify

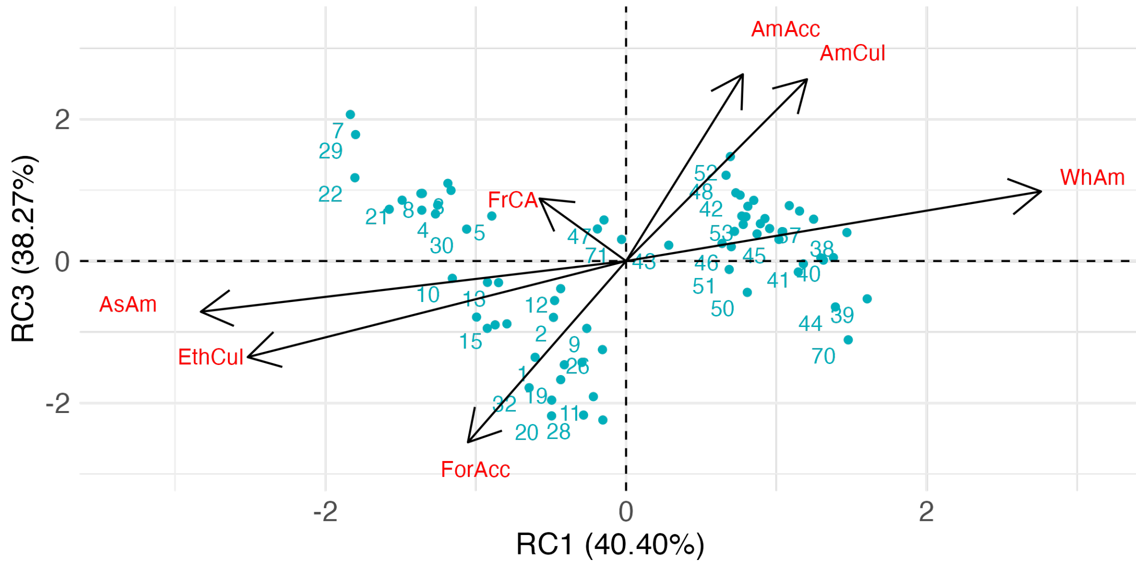

Figure 5.2: Biplot of the two key Varimax-rotated components, interpreted as representing Ethnic Background (RC1) and American Upbringing (RC3) dimensions. Individual factor scores for each photo (points numbered from 1-72) are overlaid with factor loadings for each of the seven variables (arrows).

three rough clusters of photos. To an extent, these cluster correspond to three of the quadrants: high scores on both RC1 and RC3 (upper right), low scores on RC1 but high scores on RC3 (upper left), and low scores on both RC1 and RC3 (lower left).

The larger group of photos on the right side of the plot are characterized by relatively high scores on Ethnic Background (RC1), indicating photos with a clear White American perception. The majority of these photos were also close to or above zero on American Upbringing scores (RC3), indicating that most White American-perceived photos in the sample were also perceived as having grown up in the US. At the same time, some of the White American-interpreted photos fall outside the tighter cluster, scoring moderately low on American Upbringing (RC3) which indicates that certain photos were seen as less clearly American by respondents.

On the left side of the plot, there is a more diffused set of photos that could potentially be described as falling into two loose clusters, with some photos landing in between. These are overall characterized by lower than average scores on Ethnic Background, indicating photos more likely perceived as Asian American (and engaged in ethnic culture) than White American. Specifically, the photos in the upper left quadrant tend to have the lowest Ethnic Background (RC1) scores, suggesting impressions most closely tied to being Asian American, while also having high American Upbringing (RC3) scores. Together, this indicates that a subset of the photo subjects looked specifically like American Americans who grew up in the US, speaking with an American accent (not a foreign accent) and participating in *both* American and ethnic culture.

Separately, a group of photos seem to cluster together in the lower left quadrant with clearly low American Upbringing (RC3) scores, together with moderate-to-low Ethnic Background scores. This indicates a clear impression of speaking with a foreign accent and *not* participating in American culture, but a somewhat weaker association with Asian American/ethnic cultural impressions than the previous group; this pattern may be due to these photos being less likely to be interpreted as American, and thus less likely to be labelled "Asian American", even if they were interpreted as Asian. This profile thus represents Asian-perceived photo subjects who looked like they were less likely American, both speaking with a foreign accent and being aligned with ethnic culture only.

A few photos also fall in the middle of the four quadrants, suggesting disagreement in their social impressions on both Ethnic Background and American Upbringing dimensions. Overall, however, the results of the study confirm that visual elements of a photo picturing a face can lead Californian Asian Americans to evaluate them with different social characteristics, including expectations of linguistic style.

5.3.3 Interim Discussion

The results show that individuals depicted in photos can be differentiated along axes relating to Asian American ethnic background and acculturation. People's visual styles did elicit impressions of different AAC social types that could be interpreted as ethnic-oriented or mainstream-oriented Asians. Specifically, photos that scored high on the American Upbringing component could be considered to be more mainstream and those that scored low could be considered more ethnic.

Notably, though, while the two Culture variables were intended to assess cultural alignment in a similar way, perceived participation in ethnic culture was found to pattern with the race/ethnicity variables instead. At least for this sample of photos intended to represent Asian American and White American women, impressions of whether someone engages in ethnic cultural activities seems to be tied to their racial/ethnic background regardless of where they grew up. One potential reason for this may be due to the particular range of photos included in the study, such that all the Asian American-perceived photos also happened to suggest likely ethnic cultural engagement or biculturalism as well. Alternatively, Asian Americans may tend to make a blanket assumption that other Asian Americans are more likely bicultural than monoculturally American and therefore participate in ethnic activities by default.

Regardless, these results mean that Asian American-perceived photos in this sample that were considered more mainstream were interpreted as specifically *bicultural*—they were seen as involved in both ethnic and American culture. Photo subjects were not typically read as Asians who were aligned with American culture only, which might be more similar to a Whitewashed Asian social type. In the next survey, I narrow in on selected photos using the composite variable for American Upbringing, asking detailed questions about what social impressions Californian Asian Americans get from these visually-cued social styles.

5.4 Survey 1b: Photo-based Impressions

Following up on the first survey, which confirmed that Californian Asian Americans can and do make social judgments of AACs based on visual styles along a continuum of acculturation, this survey aimed to describe at a more detailed level *how* certain AAC-interpreted photos—designated as representing different social types—may be socially evaluated on a range of characteristics. Through this, we can examine the social and linguistic associations that comprise ideologies of ethnic- and mainstream-oriented AACs. In preparation for the subsequent experiments, this task served to provide fine-grained sociolinguistic information for each potential visual stimulus photo, allowing the most representative and well-matched photos to be selected as examples of each target persona.

To provide social attribute assessments of each photo, a photo evaluation task was employed. Participants were shown a photo, then asked a set of open-ended and closed-ended questions about their impressions of the photo subject's social characteristics. This included questions about race/ethnicity and cultural alignment, but also questions asking about age, occupation, personality traits, and how they speak. By assessing the responses for different groups of photos based on Survey 1a results, we can describe both specific and potentially generalizable social impressions with regards to perceptions of being a mainstream-oriented or ethnically-oriented AAC.

5.4.1 Methods

5.4.1.1 Visual Stimuli

The visual stimuli consisted of 30 photos depicting a person's face.[4] Images were a subset of the 72 photos presented in Survey 1a. Ten photos each were selected to represent three main types of people or personae, labeled as Ethnic Asian (EA), Mainstream Asian (MA), and Mainstream White (MW). The process for selecting photos for each group was based on the responses from Survey 1a.

First, the photos were filtered only to those selected over 50% of the time when described as "from California". This was so that photos would all be considered plausibly Californian, matching the participant pool in regional location. Second, within each condition, photos were filtered to those that were selected over 50% of the time for the relevant race/ethnicity label (i.e., "White American" for Mainstream White and "Asian American" for Ethnic/Mainstream Asian).

Then, the narrowed demographic subsets were further refined to represent each persona. This was done by filtering photos based on the emergent composite scores acquired from the rotated PCA, namely the Ethnic Background (RC1) and American Upbringing (RC3) components. To be included in the Mainstream White group, photos must have scored above the mean (i.e., zero) on both Ethnic Background and American Upbringing. The criteria was the same for the Mainstream Asian group except photos needed to score *below* the mean on Ethnic Background. Finally, inclusion in the Ethnic Asian group required scoring below zero on both Ethnic Background and American Upbringing.

In the last step, photos were ranked by American Upbringing scores and the top ten photos were selected for inclusion. For Mainstream personae, photos were ranked starting from the highest score (i.e., most American-raised) while for the Ethnic persona, photos were ranked from the lowest score (i.e., least American-

[4]Stimuli are available from the author on request.

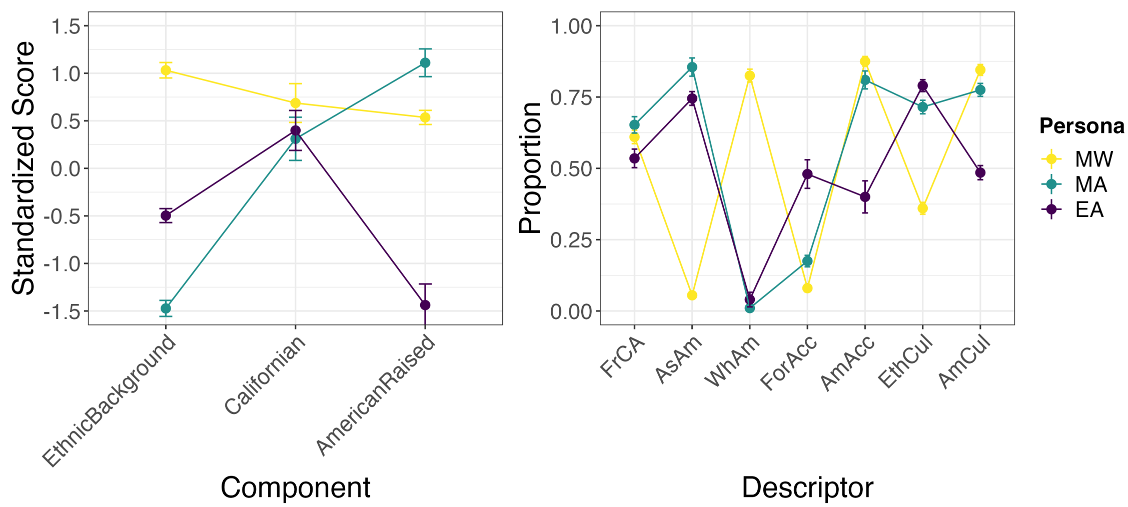

Figure 5.3: Parallel coordinates plot showing the mean values (error bars represent standard error) of the component scores (left) and descriptor proportions (right) per visual stimuli persona photo group.

raised). At this stage, photos were reviewed and a small number were manually excluded. Reasons for exclusion included that the photo was recognized as a stock photo or the photo looked visually incoherent with the other selected photos (e.g., cropped too close). If a photo in the top ten was excluded, the next highest ranking photo replaced it.[5]

While this procedure resulted in at least ten options for the two Mainstream personae, only seven photos were identified for the Ethnic Asian group. In this case, all seven photos were selected first. Then, I conducted the same filtering process but expanded the criteria to include those who were interpreted as "from California" at least 40% of the time. The remaining photos for the Ethnic Asian persona were selected from the top-ranking (i.e., least American-raised) options of this subset. Figure 5.3 presents the mean component scores and per-descriptor

[5]The photo selection process was conducted based on a previous version of the Varimax-rotated PCA where there was an error such that eigenvectors (i.e., unscaled loadings) were rotated instead of the factor loadings. The final analysis, reported in Section 5.3.2, corrected this error. If this photo selection process used the corrected analysis, one photo in the Mainstream Asian group would have been different: Instead of the tenth-ranked photo being included, the current selections include the eleventh-ranked photo.

proportions for each of the final personae groups, representing ten photos each. This illustrates that the three sets of photos differed as intended on the various social evaluation measures from Survey 1a.

5.4.1.2 Survey & Procedure

The survey was hosted on Qualtrics. Before taking part, participants were told that the study was about social perceptions of people based on style and that they would be asked to "look at photos of people and answer questions about your impressions of the speaker's personal characteristics". The main portion of the survey was the photo evaluation task. Participants were given the following instructions:

> In this task, you will look at photos of people and describe your impressions of them. For each person, you will first see a photo and be asked to describe them in three words. Then, you will be asked to comment on or rate your impressions on various social dimensions based on how they look.

In order to reduce the length of the survey, each participant was presented with a subset of ten photos that were randomly selected from the full set of 30 photos. While the total number of photos was fixed, the number of photos per personae was variable across participant.

The task was self-paced. For each photo block, a series of 13 questions probed various social characteristics of the person in the photo (for the full questionnaire, see Appendix A; for similar social evaluation survey designs, see Campbell-Kibler, 2007; D'Onofrio, 2019; Drager et al., 2021; Leigh, 2021). These were presented across eight pages. On each page, the photo was displayed at the top of each page and one or multiple questions were displayed below. The main content questions were generally structured such that open-ended questions were ordered before rating questions; these are described in more detail below. In addition, participants were asked if they recognized the person in the photo and if anything else about the person stood out to them. No photo subjects were recognized.

At the start of each block, participants were asked to provide adjectives to describe the pictured individual, in order to obtain general impressions of visual style. Specifically the instructions were as follows: "Please describe your impressions of this person using three words. You can use whatever adjectives you like, but try not to focus just on physical attributes, or what the person is wearing." Participants entered each descriptor into a text box.

This was followed by targeted questions about demographic or personal characteristic impressions. To assess perceived age, participants were asked to select how old they thought the person looked from four options—Teens, 20s, 30s, and Over 40—where multiple selections were possible. Then, five open-ended questions asked participants to respond with their impressions about the pictured individuals' occupation, hobbies, regional origin (i.e., where they grew up), ethnic background, and speech style.

Next, participants were asked to rate the pictured individuals on sixteen 7-point semantic differential scales (see Table 5.6). These scales were grouped according to the topic or category, consisting of six scales relating to perceived stylistic expression (Style), six scales relating to perceived personal traits (Trait), and four scales relating to perceived cultural/linguistic expression (Culture). Scale groups were presented in a fixed order across three pages.

Within each category, scales were presented in a random order while scale direction was uniform. This meant that the reference labels (e.g., Intelligent) each page were always on the same side (e.g., left side) of each scale while the opposite or negated counterparts (e.g., Unintelligent) were always on the other side (e.g., right side). However, across blocks, the direction of scales alternated and this was further counterbalanced across participants. In other words, half of the participants were presented with reference-to-negated scales on the first page (e.g., Cool-Uncool) and would be shown negated-to-reference scales on the second page (e.g., Not confident-Confident), then reference-to-negated scales again on the third page (e.g., American-Not American). The other half saw the opposite pattern. This design was intended reduce the potential effect of direction on responses while providing some consistency for the respondent within a set of ratings. In addition, because many of the reference labels had a positive valence while the opposite label had a negative valence, direction was switched between blocks in an effort to reduce the potential attribution of negative valence to scales about non-American culture (e.g., Foreign-accented); these would have otherwise been aligned with the negated labels only.

Prior to the first block, participants were provided with an example photo with modelled answers to the description question (i.e., three adjectives), as well as a preview of all the questions they would be asked per photo. The example adjectives were intended to help communicate the types of answers that were expected given the open-ended nature of the task. Exposing participants to all of the questions ahead of time also helped to ensure that all photos were judged with the same expectations going in; this included ensuring that all participants were exposed to the adjectives used in the semantic differential scales before performing the open-ended description for the first photo.

After completing the evaluation task, participants were asked what they thought the task was about, what they thought about the task as they were completing it, and whether they noticed anything else about the stimuli or task. Then, they answered questions about their demographic background, including age, gender, and ethnicity, as well as questions about their social network with regards to various race/ethnicity groups. The survey took on average 40 minutes to complete and each participant who did so received $7 in compensation.

5.4.1.3 Participants

Sixty Asian Americans from California completed the study remotely via Prolific (www.prolific.com). The same screening criteria as Survey 1a was used. Of this sample, fourteen had previously completed Survey 1a.

The mean age was 32 (M=31.75, SD=11.08, range=18-63). The sample included 23 women, 34 men, and 2 nonbinary individuals (1 chose not to respond). All were living in California at the time, and the vast majority were born there (1 was born in Asia, 3 were born in another US state). Further, all participants self-identified as Asian (38 East Asian, 20 Southeast Asian, 2 East and Southeast Asian), including four who identified as Mixed (see Table 5.4 for specific ethnicities; 1 did not provide a specific ethnicity). The most common ethnic backgrounds represented in this sample were again Chinese, Vietnamese, and Filipino. To assess social network exposure, participants provided estimates of the proportion of people they interacted with as close contacts, regular contacts, and the broader community

Ethnicity	**Total (n=60)**	**Man (n=34)**	**Woman (n=23)**	**NB (n=2)**	**NA (n=1)**
Chinese	**21**	13	7	0	1
Vietnamese	**10**	7	3	0	0
Filipino	**8**	4	3	1	0
Japanese	**5**	2	3	0	0
Korean	**5**	3	2	0	0
Cantonese/Hongkongese	**4**	1	2	1	0
Hmong	**2**	1	1	0	0
Filipino-Italian	**1**	0	1	0	0
Japanese-European	**1**	0	1	0	0
Taiwanese	**1**	1	0	0	0
Thai	**1**	1	0	0	0
NA	**1**	1	0	0	0

Table 5.4: Number of participants by self-identified ethnicity and gender.

who would fall into different racial/ethnic groups. On average, the amount of social experience with other East and Southeast Asians could be described as moderately high, but there was much variation across participants (M=0.63, SD=0.30, range=0.04-1).

5.4.1.4 Data Analysis

Each of the 30 photos (10 EA, 10 MA, 10 MW) received 20 sets of evaluation responses from different participants. In these results, I focus on the key questions relevant to characterizing the target personae, namely the three open-ended responses about ethnic background, regional origin, and speech style, as well as the 16 semantic differential scale ratings.

For each of the open-ended questions, 600 text descriptions were collected across photos and participants. In order to provide quantified estimates of the overall social impressions per group, responses were coded systematically as follows. First, ethnic background responses were coded as Asian (e.g., Chinese, Taiwanese, Filipino/a, Malaysian) or White (e.g., Caucasian, Anglo, Finnish, German), based on ethnic groups associated with East and Southeast Asian countries or European countries as a rough approximation. If only "American" was offered as the ethnic background, it was coded as such. Any other responses were categorized as Other (e.g., Hispanic/Latina, Armenian, Pacific Islander). If multiple options were mentioned, responses were classified into the first applicable category in this order: Asian, White, American.

Second, geographic locations were extracted from each regional origin response, then coded for country and continent (North America, Asia, Europe, South America); if no specific location was identified (e.g., responses referencing "suburbs" and "cities" rather than naming a city), the response was coded as Other. If multiple options were mentioned, responses were classified into the first applicable category in this order: North American, Asia, Europe, South America.

Third, responses about speech style were coded for references to an accent or (non-)native English. One set of responses was grouped under the label of Foreign Accent, intended to assess impressions of accented English linked to Asian faces. This included responses mentioning an unspecified accent, foreign accent, Asian or Pacific Islander-related accent, or non-fluent English. A second category was labeled Mainstream Accent, capturing ideologies about sounding “neutral” or American to Californian respondents. Responses coded as Mainstream included phrases referring to a lack of accent, a standard, general American, or Californian accent, and “native”-sounding English. A third category, Other Accent, grouped together mentions of accent that specifically named other regional American or European accents (e.g., British accent, country accent, Southern drawl). All other responses were coded as Other. Responses were first classified as Other Accent if applicable, followed by Foreign Accent, then Mainstream Accent.

In terms of the rating responses, 9589 ratings were collected in total across scales, photos, and participants (11 were null responses with no rating provided). Because semantic differential scales varied in direction by category and participant, ratings were first normalized such that the maximum value of 7 always corresponded to the reference label (e.g., Enthusiastic) while the minimum value corresponded to the opposite or negated label (e.g., Unenthusiastic).

To assess whether and how photos associated with each persona (as categorized by responses from Survey 1a) differed on the detailed social evaluation ratings (provided by participants in the current survey), a one-way MANOVA was conducted in R. The independent variable was PERSONA and the dependent variables were the ratings on the sixteen scales. Nine rows containing a missing response for any variable were removed prior to analysis. Univariate ANOVAs were further conducted to examine which rating scales differed between PERSONA photo groups. Bonferroni correction for multiple comparisons was applied to interpret the results across the 16 tests and Tukey HSD post-hoc pairwise comparisons were conducted to describe which groups differed on each relevant scale.

5.4.2 Results

5.4.2.1 By-Voice Text Responses

To provide a sense of how each persona was interpreted in terms of demographics, I first describe the results of the targeted open-ended questions. The proportions of each response code per PERSONA photo group are presented in Table 5.5. In terms of perceived ethnic background, EA and MA photos were nearly always identified as Asian, while MW photos were mostly identified as White (or alternatively, American). When considering the most common response per photo, the results were unanimous within each persona: All ten EA and MA photos were most likely perceived as Asian and all ten MW photos were most likely perceived as White.

For regional origin, EA photos were overall variably identified as growing up in either North America or Asia. The most common response by photo was evenly split such that five EA photos elicited more North America responses while the other five EA photos elicited more Asian location responses. On the other hand, both MA and MW photos were mainly identified as growing up in USA or Canada. Where they differed was that non-North American responses were mostly Asian locations for MA photos but European locations for MW photos.

The types of guesses provided for speech style were highly varied. With regards to the subset of responses relating to accentedness, there was a difference between EA on the one hand and MA and MW on the other. Specifically, both MA and MW photos received the most Mainstream Accent comments, including mentions of speaking with no accent or a general American/Californian accent.

Question	Response Code	EA	MA	MW
Ethnic Background	Asian	0.98	0.95	0.00
	White	0.00	0.01	0.83
	American	0.01	0.01	0.12
	Other	0.01	0.04	0.05
Regional Origin	Asia	0.47	0.12	0.00
	North America	0.46	0.80	0.83
	Europe	0.01	0.00	0.12
	South America	0.00	0.00	0.01
	Other	0.06	0.08	0.05
Speech Style	Foreign Accent	0.24	0.09	0.05
	Mainstream Accent	0.14	0.23	0.24
	Other Accent	0.02	0.01	0.07
	Other	0.59	0.68	0.64

Table 5.5: Proportions of coded responses for open-ended questions about ethnic background, regional origin, and speech style, averaged across all responses for each of ten photos per PERSONA condition (EA=Ethnic Asian, MA=Mainstream Asian, MW=Mainstream White). For ease of visual interpretation, values above 0.2 are shaded in grey.

EA photos, on the other hand, received more Foreign Accent comments, including mentions of unspecified or Asian/Pacific Islander accents.

Altogether, the two Mainstream personae were generally interpreted as having grown up in North America, but MA was perceived as ethnically Asian while MW was perceived as White. They were also often expected to speak with a native or mainstream American English accent. In contrast, the Ethnic Asian persona was interpreted as Asians who more likely grew up in Asia and spoke with a foreign or Asian-associated accent in English. In general, these open-ended responses show that the photo subjects in each group were interpreted as aligning with the expected characteristics of each persona, including how they might speak.

5.4.2.2 By-Group Ratings

Next, I report on the results of the semantic differential ratings targeting various aspects of cultural alignment, personal traits, and stylistic expression in order to assess how each persona was socially evaluated in detail. The mean rating values for each photo group are displayed in Table 5.6 and Figure 5.4.

Overall, the Culture scales appear to most differentiate the three sets of photos while the Trait and Style categories were on average quite similar between the

Category	Rating Scale	EA	MA	MW
Culture		3.73 (1.96)	5.26 (1.72)	5.93 (1.66)
	American–Not American	3.96 (1.93)	5.38 (1.68)	5.89 (1.73)
	Native speaker of English–Non-native speaker of English	3.79 (2.01)	5.50 (1.70)	6.06 (1.58)
	American-accented–Foreign-accented	3.68 (1.98)	5.39 (1.67)	5.80 (1.78)
	Aligned with American culture–Aligned with ethnic culture	3.52 (1.90)	4.75 (1.73)	5.96 (1.56)
Trait		5.46 (1.26)	5.51 (1.33)	5.46 (1.26)
	Enthusiastic–Not enthusiastic	5.26 (1.29)	5.50 (1.33)	5.57 (1.21)
	Confident–Not confident	5.38 (1.33)	5.56 (1.35)	5.66 (1.15)
	Friendly–Unfriendly	5.58 (1.22)	5.70 (1.33)	5.56 (1.28)
	Likable–Unlikable	5.58 (1.17)	5.62 (1.26)	5.46 (1.29)
	Attractive–Unattractive	5.35 (1.35)	5.15 (1.32)	5.40 (1.27)
	Intelligent–Unintelligent	5.61 (1.19)	5.53 (1.34)	5.13 (1.30)
Style		4.63 (1.54)	4.75 (1.71)	4.73 (1.61)
	Feminine–Not feminine	5.77 (1.10)	5.11 (1.57)	5.65 (1.27)
	Clear speaker–Unclear speaker	4.92 (1.40)	5.44 (1.33)	5.62 (1.16)
	Cool–Uncool	4.84 (1.34)	5.09 (1.47)	5.01 (1.23)
	Casual–Formal	4.04 (1.68)	5.16 (1.73)	4.82 (1.65)
	Nerdy–Not nerdy	4.19 (1.56)	4.07 (1.80)	3.52 (1.61)
	Slow speaker–Fast speaker	4.02 (1.34)	3.62 (1.52)	3.75 (1.32)

Table 5.6: Mean ratings (SD) on sixteen 7-point semantic differential scales, as well as mean ratings of Culture, Trait, and Style categories, averaged across all responses for the ten photos per PERSONA condition (EA=Ethnic Asian, MA=Mainstream Asian, MW=Mainstream White). Scale labels represent the interpretation of the maximum value of 7 (reference term on the left) and the minimum value of 1 (opposite term on the right).

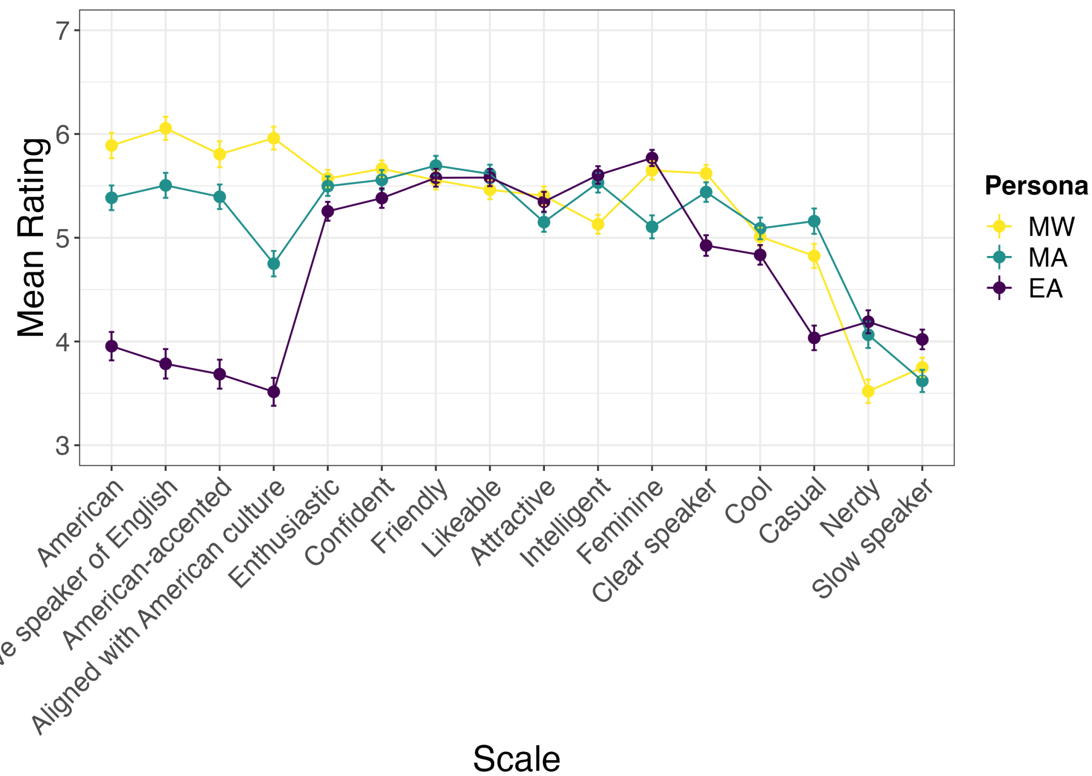

Figure 5.4: Parallel coordinates plot showing the mean values (error bars represent standard error) of the semantic differential scale ratings per PERSONA photo condition.

groups. This pattern is clearly illustrated in Figure 5.4. There are relatively large average differences between the three PERSONA groups on the four Culture scales, especially between EA on the low end and both MA and MW on the higher end.

While the differences are relatively similar for three of the scales, there seems to be a more equally distributed three-way distinction between EA, MA, and MW on ALIGNED WITH AMERICAN CULTURE. EA was the least aligned with American culture (or most aligned with ethnic culture), MW was the most aligned with American culture, and MA fell directly in between. Most of the other scale ratings were quite similar between the three groups (e.g., FRIENDLY) though a few scales appear to differ in ratings between one specific PERSONA group compared to the other two (e.g., FEMININE).

The results of the one-way MANOVA support these observations. At least two of the PERSONA photo groups were found to be different when accounting for all 16 scales (F[2,588]=11.139, p$<$ 0.001; Wilk's λ =0.5818, partial η^2 = 0.23). In other words, photos selected to represent different PERSONA along dimensions of race/ethnicity and acculturation were able to be differentiated by some combination of social evaluation ratings.

Post-hoc univariate ANOVAs further indicate that groups differed on nine scales overall. These include all four Culture scales, namely AMERICAN (F[2,588]= 62.538, $p<$ 0.001), NATIVE SPEAKER OF ENGLISH (F[2,588]=85.868, $p<$ 0.001), AMERICAN-ACCENTED (F[2,588]=74.416, $p<$ 0.001), and ALIGNED WITH AMERICAN CULTURE (F[2,588]=96.362, $p<$ 0.001). Pairwise comparisons showed that the EA group differed above the threshold of statistical significance from both MA and MW PERSONA groups on AMERICAN (p<0.001), NATIVE SPEAKER OF ENGLISH (p<0.001), and AMERICAN-ACCENTED (p<0.001) ratings. MA and MW did not significantly differ from each other on these same scales: AMERICAN (p=0.021), NATIVE SPEAKER OF ENGLISH (p=0.009), and AMERICAN-ACCENTED (p=0.090). On the other hand, all three groups were found to significantly differ on ALIGNED WITH AMERICAN CULTURE (p<0.001).

Together, this indicates that respondents in this survey distinguished between the three groups of photos based on perceived alignment with American or ethnic culture. The EA persona was seen as most ethnically-aligned, the MA persona was seen as not clearly aligning with either ethnic or American culture, and the MW persona was seen as most aligned with American culture. The three other culture-related scales appear to pattern with the dimension of acculturation—from mainstream to ethnic—which confirm that both MA and MW photo subjects were seen as generally likely to be American or American-raised, including being a native speaker of English and American-accented.

Five other scales were flagged as relevant to differences between PERSONA conditions. INTELLIGENT was the one Trait scale that was identified as different between groups (F[2,588]=7.3227, $p<$ 0.001). In addition, four of the Style scales were found to differ: FEMININE (F[2,588]=13.798, $p<$ 0.001), CLEAR SPEAKER (F[2,588]= 15.462, p< 0.001), CASUAL (F[2,588]=23.38, p< 0.001), and NERDY (F[2,588]=8.4934, p< 0.001). With conservative thresholds for statistical significance, pairwise comparisons highlight three patterns of PERSONA-based differences within these scales. First, EA differed from MW on INTELLIGENT (p=0.001) and NERDY (p<0.001) ratings. Second, EA differed from both MA and MW on CLEAR SPEAKER (p<0.001) and CASUAL ratings (p<0.001). Third, MA differed from both EA and MW on FEMININE ratings (p<0.001).

These results may provide additional context for how each of the personae are conceptualized as an ideological social construct. The first pattern signals that respondents associated the EA persona photos with characteristics of intelligence and nerdiness, more so than the MW persona photos. Although model results do not support the conclusion that MA photos differ from MW on these scales, MA photos numerically pattern closely with EA on average. The second pattern indicates ways that the EA persona photos may differ from the two mainstream personae photos. EA photos elicited impressions of speaking less clearly while also being read as stylistically more formal than casual. Finally, the third pattern suggests that MA persona photos were seen as portraying a less feminine style than the other two persona.

In general, these results demonstrate that Californian Asian Americans have different ideological constructs for Asian American who are seen as ethnically-oriented and those who are seen as bicultural between ethnic and American cultures. These both differ from the idea of a mainstream-oriented white American. In addition, a few elements—like being American, American-accented, and a clear speaker—that pattern together for more mainstream-oriented Asian and white

Americans in contrast to ethnically-oriented Asian Americans.

5.4.3 Interim Discussion

These results both confirm and extend the findings of Survey 1a. Californian Asian American can and do distinguish between an ethnically-oriented Asian American persona and a mainstream-oriented bicultural Asian American persona. Moreover, the latter is perceived as more similar to a mainstream-oriented White American persona on various social dimensions, including those relating to being American and those that are potentially ideologically-linked, like speaking clearly and portraying a casual style.

While the overall conclusions remain the same, the current findings provide more nuance than Survey 1a in a few ways. For example, the open-ended responses demonstrated that the two Asian personae were both equally interpreted as ethnically Asian, but the key difference was that the mainstream persona looked like they grew up in North America. This clarifies exactly how the two groups of photos differ, since the descriptor of "Asian American" from Survey 1a likely led to responses that conflated ethnicity and regional origin.

Plus, rather than simply linking accent to cultural alignment as the results of Survey 1a suggested, the current results are better able to separate perceptions of someone who grew up or lives in America and someone who participates in American culture. The former in particular appears to be linked to ideologies about how someone speaks English. That is, being American, regardless of race or ethnicity, seems to be closely tied in ideology to being a native speaker, American-accented, and a clear speaker. On the other hand, cultural participation, appears to be somewhat independent of how one sounds—at least for mainstream-oriented Asian Americans who were rated as relatively less aligned with American culture and relatively more American-accented.

Other differences in terms of personal traits and style impressions offer some potential insights into the ideological space that each persona is embedded in. The fact that photo subjects interpreted as ethnically-oriented Asian Americans were linked to being more intelligent and nerdy calls to mind Asian cultural stereotypes of being particularly studious and academic-focused. Mainstream-oriented photos being rated as more casual—and possibly mainstream-oriented Asian photos being rated as less feminine—could be connected to perceived differences in appearance-related and behavioral norms between cultures. I explore these connections further in the General Discussion (Section 5.6).

Overall, the results indicate that, for Asian Americans in California, the visual style of Asian-perceived individuals is tied to linguistic style, at least at the level of accentedness. That is, the way Asian Americans look, including how they dress or style their hair, can lead to expectations about whether they speak English with a perceived mainstream accent or foreign accent. To further explore ideologies about how mainstream-oriented and ethnic-oriented Asian Americans may be expected to speak similarly or differently from each other, the next survey includes auditory stimuli to examine how particular voices and/or speech styles may be interpreted as belonging to different AAC social types.

5.5 Survey 2: Voice-based Impressions

The final survey incorporates both recorded voices and photo. It follows a similar format to Survey 1b in order to describe how certain voices may be socially evaluated in a detailed manner, as well as includes a photo selection task similar to Survey 1a where respondents are prompted by voices rather than descriptors. These results provide insight into whether and how speech—in the absence of potentially identifiable phonetic features like /oʊ/-backing and /ð/-stopping—may be linked to ideologies about AAC social types. They also provide context for how these voices may be interpreted for the experimental phase of this research.

In the speaker evaluation task, participants first listened to a short recording of a voice reading a series of words. Then, they answered a set of open-ended and closed-ended questions about their impressions of the person's social characteristics. After providing responses for all voices, they heard each of the recordings again and were asked to select faces that they felt matched with each voice. This approach provided multiple ways to assess how each voice may be socially interpreted at the level of macrosocial identities as well as at the level of microsocial personae that are differentiated by degree of acculturation.

5.5.1 Methods

5.5.1.1 Auditory Stimuli

The auditory stimuli consisted of 11 speech sample recordings of words read by self-identified Asian American/Canadian women.[6] Each speech sample contained the same 12 words in the same order spoken by a single voice. The isolated word recordings included in these speech samples were recorded as part of the full set of recordings used to create auditory stimuli for the speaker identification experiment reported in Chapter VI (see Appendix C for the full wordlist). Here, I describe only the key details relevant to this survey. For a detailed description of the stimuli design and recording procedure, see Section 6.3.1.

The 11 speakers were self-identified Asian American/Canadian women between the ages of 23 and 42 (M=29.7). Two of these speakers identified as mixed or biracial as well. Aside from two speakers who primarily described themselves as Asian American, all speakers described themselves with East Asian ethnic identities (4 Chinese, 3 Korean, 2 Hong Kong Chinese/Cantonese). In terms of where

[6] Stimuli for 10 speakers are available from the author on request; one did not consent to their recordings being shared.

speakers spent their time before age 18, four had grown up mainly in the US, five had grown up mainly in Canada, and two had grown up mainly in Asia[7]. Each speaker self-recorded the full word list (Appendix C) in a sound-attenuated booth (n=5) or quiet room (n=6). Recording equipment varied between speakers, including use of an external microphone (n=6), laptop microphone (n=4), or handheld recorder (n=1).

Words selected for this task were designed as filler words for the speaker identification experiment (see Chapter VI). These involved one of four variables associated with the California Vowel Shift (e.g., Wolfram & Schilling, 2015) or non-regional variation (e.g., Eckert, 2008a): /u/-fronting (e.g., *moody*), /æN/-raising (e.g., *channeled*), /æ/-backing (e.g., *padding*) and coda cluster reduction or /t/-deletion (e.g., *fastball*). Crucially, they do not contain sounds involved in the critical AAC-associated variables (/oʊ/-backing and /ð/-stopping). In addition, the words in this task were produced without the target variation, namely /u/ produced as a (centralized) high back vowel, /æ/ produced as a low front vowel, and the coda cluster produced without deletion. This enabled examination of how each voice may be holistically evaluated in the absence of socially marked segmental features that will be manipulated in the speaker identification experiment.[8]

The full wordlist contained ten words per variable. For the current task, three words were randomly selected for each of the four variables. Then, the twelve selected words were randomly shuffled. The final ordered word list was as follows: "tattletales, restlessness, moody, chattering, soothingly, padding, randomness, scooted, channeled, banditry, wistfulness, fastball."

[7]One of these speakers moved from Asia to the US at age 13. The other speaker grew up in Asia speaking English with family members and attended an international school.

[8]Note that keeping certain features constant was only intended to provide some degree of control across voices (rather than, for example, an attempt at approximating a perceptibly "unmarked" accent). The selected features carry their own social meanings and the recordings are expected to contain other socially meaningful phonetic features aside from the targeted ones that listeners draw on to form holistic impressions.

All recordings were manually reviewed by the author and one token per word and speaker was extracted. The peak amplitude of each word recording was normalized to 60 dB. Recordings were then padded with 250 ms of silence before and after the word. Finally, the single word recordings were concatenated in Praat, resulting in 500 ms of silence between each word. The final 11 auditory samples were on average 16.2 seconds in duration (range=13.5-18.2).

5.5.1.2 Visual Stimuli

The visual stimuli consisted of 18 photos of faces, a subset of the 30 photos from Survey 1b.[9] Three photos were selected to represent six emergent social types. These included three groups of Asian-perceived photos and three groups of white-perceived photos. The Asian photos selected at this stage were intended to best represent the two targeted Asian American personae, namely Mainstream Asian (MA) and Ethnic Asian (EA) Americans. In addition, an intermediate category with elements of both targeted personae was included for consideration, labeled as the Intermediate Asian (IA) group. Separately, three white American photo groups—Mainstream White (MW), Ethnic White (EW), and Intermediate White (IW)—were formed to serve as comparisons.

Photos were divided into groups based on the Culture-related semantic differential scale ratings from Survey 1b. The four ratings were standardized by participant and converted into three variables intended to capture the relevant ways that social perceptions of acculturation varied across the photos. One variable was the average of a photo's AMERICAN, NATIVE SPEAKER OF ENGLISH, and AMERICAN-ACCENTED ratings; higher scores were interpreted as representing a stronger perception of being "American-raised" in terms of upbringing. The second variable was a photo's rating for ALIGNED WITH AMERICAN CULTURE,

[9] Stimuli are available from the author on request.

where higher scores were interpreted as representing a stronger perception of being "American-aligned" in terms of cultural orientation. The third variable was the calculated as the difference between "American-raised" and "American-aligned" scores, where *lower* scores represented a stronger perception of being "bicultural". In other words, a negative "bicultural" score signified that a photo was perceived as less "American-aligned" than they were perceived as "American-raised".

Using these variables, photos were organized into six groups of three. The choice of three photos per group was based on a combination of factors, including constraints of the stimuli needed for the following experiment guided by constraints of the photo options available. In terms of the experimental constraints, multiple photos were required per persona group and photos needed to represent Asian- and White-perceived individuals in equal numbers. Given that only ten White-perceived photos were included, the options included groups of five, four, or three. Inspection of the key acculturation-related variables across photos revealed some cases where there were clusters of three photos with similar ratings, and no other photos with similar ratings. In order to create photo groups that were as internally consistent as possible based on the key variables, I decided to select three photos per group. In addition, because some of the photos previously classified as either EA or MA based on Survey 1a overlapped on the key variables, I decided to select photos for the three Asian personae from the combined pool of all Asian-perceived photos, relying on the responses from Survey 1b alone.

To create the three Asian persona groups, Asian-perceived photos were first separated by "American-aligned" scores into groups that were perceived as relatively more or less "American-aligned". In order to operationalize this comparison, the mean ratings of Survey 1b's EA and MA photo groups on the ALIGNED WITH AMERICAN CULTURE scale were used as benchmarks. Thus, the refined MA group was formed by initially filtering for photos that had "American-aligned" scores above the Survey 1b MA group mean (i.e., the most American-aligned), then selecting the three photos with the lowest values on the "bicultural" score (i.e., the most bicultural). The same general procedure was followed to create the other two photo groups. For EA, the criteria was the opposite, requiring "American-aligned" scores below the Survey 1b EA group mean (i.e., the least American-aligned) and within that, photos with the highest "bicultural" scores (i.e., the least bicultural). For IA, photos were filtered to those with intermediate "American-aligned" scores between the Survey 1b MA and EA group means; then, just like for the MA group, the three photos with the lowest "bicultural" scores (i.e., the most bicultural) were chosen.

To create comparable White persona groups, photos classified as MW in Survey 1b were divided into groups of three based on the maximum and minimum "American-aligned" scores. First, the three photos with the highest (MW) and the lowest (EW) "American-aligned" scores were selected. Because most photos had high scores, the three photos with the next lowest "American-aligned" scores were selected to form an intermediate group (IW) that differed as much as possible from the previous two groups.

Upon review of the full set of social evaluation responses per photo, any photos I judged as outliers or distinctly different on any measure were removed in order to maintain overall consistency of social impressions within groups. These were replaced with the next best option. For example, two photos (one Asian and one White) were excluded for being described with particularly negative valence[10] compared to all other photos, and two photos were excluded for being differently racialized for all other Asian photos (e.g., perceived as less likely Asian, or perceived as Hawaiian).

[10] Positive or negative valence was assessed based on sentiment analysis of text descriptors that participants provided for each photo in Survey 1b. Averaged sentiment scores across all descriptors per photo were used to estimate valence.

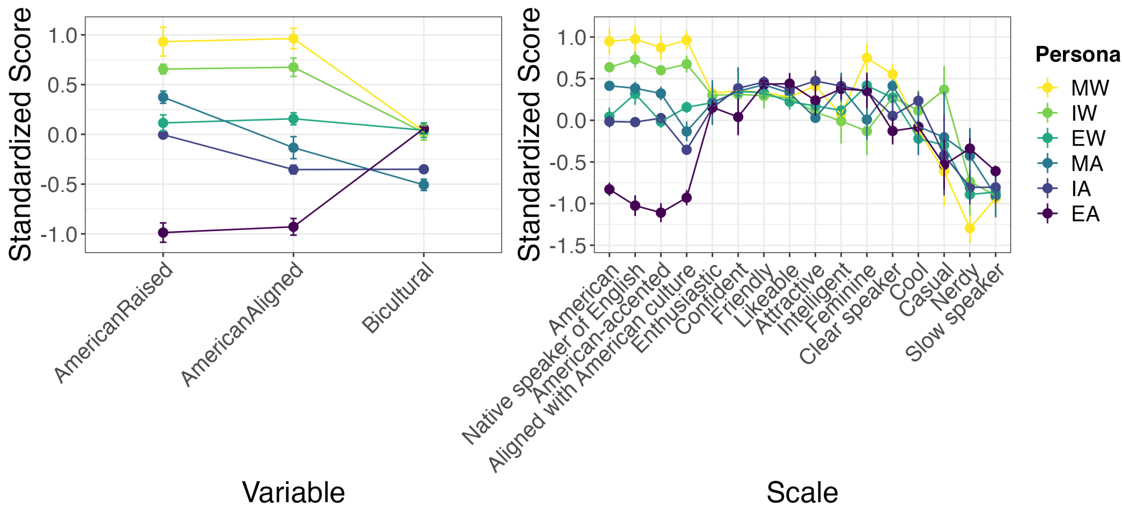

Figure 5.5: Parallel coordinates plot showing the mean values (error bars represent standard error) of the variable scores (left) and rating scores (right) per persona photo group.

Figure 5.5 displays mean scores for each of the six personae groups, representing three photos each. Overall, this demonstrates that each set of photos differed on the key acculturation-related variables. It also shows how they vary on the other social evaluation ratings from Survey 1b. Within both Asian and White categories, the Mainstream and Ethnic persona groups are maximally different from each other on "American-raised" and "American-aligned" variables, while the Intermediate persona groups are in the middle. In both cases, the values for the Intermediate persona are more similar to the corresponding Mainstream persona than the Ethnic persona, including the pattern of lower "American-aligned" and "bicultural" scores found only for MA and IA photos. Aside from differing slightly on the Culture-related scales, MW and IW appear to also differ noticeably on a few stylistic scales, including FEMININE and CASUAL; as such, the interpretation of the IW scale may be qualitatively different from the MW.

5.5.1.3 Survey & Procedure

The survey was hosted on Qualtrics. Before taking part, participants were told that the study was about social perceptions of people based on style and that they would be asked to "listen to audio recordings of people's voices and answer questions about your impressions of the speaker's personal characteristics". The task was broken up into two parts, presented as Part 1A and 1B. Out of the 11 voices, 6 voices were randomly selected per participant and presented in a random order. Voices were encountered in the same order in Part 1A and 1B.

Part 1A was a photo evaluation task in the same format as Survey 1b. Participants were given the following instructions:

> In this task, you will listen to audio recordings of people's voices and describe your impressions of them. For each person, you will first listen to an audio clip of a voice reading a list of words and you will be be asked to describe them in three words. Then, you will be asked to comment on or rate your impressions on various social dimensions based on how they sound. Each voice will be saying the same set of words so please make judgments based on how they sound, not what they are saying.

At the top of each page, an audio player was presented along with the following instructions: "Press the play button to hear the recording. Play it as many times as you like."

The questions were based on Survey 1b in order to facilitate comparison between the two sets of social evaluations (see full questionnaire in Appendix A). The majority of questions remained unchanged, other than modifying the wording to reference voices rather than photos where needed. In particular, the semantic differential scales were the same. A number of omissions and additions were made to accommodate the fact that making social judgments based on auditory voice cues (especially when the stimuli are relatively controlled recordings of the same read word list rather than naturalistic speech) is a different and likely more difficult task compared to making social judgments of people based on visual cues (especially as photos were specifically intended to vary in style).

The main addition was a set of eleven 7-point likert scale questions, placed after the semantic differential scales. These asked the participants to rate the likelihood that the person they were hearing (a) had grown up in a particular part of the world (US/Canada, Europe, Asia); (b) had grown up in a particular region of North America (California, the Midwest, Southern US, elsewhere in the US or in Canada); and (c) had a particular racial/ethnic background (Asian, Black, Latin/Hispanic, White). The other major changes were those made to the open-ended questions, including removal of the questions about occupation and hobbies as well as modification of the question about expected speech style to a question about describing the speech style or voice. In addition, they were not asked if they recognized the voices. The questions were presented across nine pages and the task was self-paced.

In Part 1B, participants were asked to connect voices to faces in order to assess the potential link between certain voices and certain personae. They were given the following instructions:

> In this next part, you will listen to the same audio recordings of people's voices, then select photos of faces that you feel match best with each voice.

The audio player of the relevant recording was placed at the top of each page along with the following prompt: "Out of these photos, is there a person or people you think this voice would match with best?". Each page displayed six photos at a time, in two rows of three. Multiple selections were possible, meaning participants could select all or none of the photos on a page.

Each of the six photos represented a different potential persona group (EA, IA, MA, EW, IW, MW). Only one set of six photos, and therefore one example of each persona, was presented per voice. The specific photos viewed by a participant for each voice was selected in a pseudrandom manner. Twelve lists of six photos were generated randomly, and one of these twelve lists was randomly selected for presentation. The six photos were randomly ordered on the page.

In more detail, the procedure to generate the twelve lists was as follows. The full set of 18 photos was first shuffled randomly to create four lists. Recall that these 18 photos were categorized into six persona groups made up of three photos each. Based on the random order of each list, the first photo representing each persona was selected to form a subset list. This was repeated for the second and third photos per persona. This process created three random subset lists of six photos from each of the four original lists, each containing one photo per persona. In this way, each photo was evenly distributed across the twelve lists, appearing four times each.

No example or question preview was provided in this survey as all participants were recruited from those who had completed Survey 1b, and had therefore seen the relevant questions before. After completing both parts of the social evaluation task, participants were asked what they thought the task was about, what they thought about the task as they were completing it, and whether they noticed anything else about the stimuli or task. Then, they answered questions about their demographic background and social network. The survey took on average 25 minutes to complete and each participant who did so received $7 in compensation.

Ethnicity	Total (n=33)	Man (n=21)	Woman (n=11)	NB (n=1)
Chinese	11	8	3	0
Vietnamese	7	5	2	0
Japanese	4	2	2	0
Filipino	3	0	2	1
Asian	2	2	0	0
Cantonese/Hongkongese	1	1	0	0
Hmong	1	1	0	0
Korean	1	0	1	0
United States	1	0	1	0
Vietnamese-Chinese	1	1	0	0
NA	1	1	0	0

Table 5.7: Number of participants by self-identified ethnicity and gender.

5.5.1.4 Participants

Thirty-three Asian Americans from California completed the study remotely via Prolific (www.prolific.com). The screening criteria was the same as the previous two surveys. This sample was specifically recruited from the participants who had previously completed Survey 1b. In addition, nine of them had also completed Survey 1a.

The mean age was 34 (M=34.15, SD=11.30, range=21-64). The sample included 11 women, 21 men, and one nonbinary individual. All participants had been born in California and currently lived there. All additionally self-identified as Asian (22 East Asian, 10 Southeast Asian, 1 East and Southeast Asian), including two who identified as Mixed (see Table 5.7 for specific ethnicities; 1 did not provide a specific ethnicity). The most common ethnic backgrounds represented in this sample were Chinese, Vietnamese, and Japanese. Social network exposure to other East and Southeast Asians was similar to Survey 1b: moderately high on average with variability across participants (M=0.62, SD=0.27, range=0.17-1).

5.5.1.5 Data Analysis

Each voice received 18 sets of evaluation responses from different participants. As for Survey 1b, I focus on the questions most relevant to characterizing the target personae, namely the three open-ended responses about ethnic background, regional origin, and speech style; the 11 Likert scale ratings; and the 16 semantic differential scale ratings.

For each of the open-ended questions, 198 text responses were collected across voices and participants. Responses were coded systematically using the same process as Survey 1b (see Section 5.4.1.4). In terms of the 7-point semantic differential rating responses, 3166 ratings were collected in total across scales, voices, and participants (2 were null responses with no rating provided), and these were again normalized for scale direction. In addition, 2160 Likert scale ratings were collected across scales, voices, and participants (18 were null responses with no rating provided).

To assess whether particular personae were more often selected in response to particular voices, I calculated the proportion of selections made for each persona and voice. In total, 1188 binary selection responses were analyzed. First, photos were labeled by their persona category. Then, the total number of times a persona category was selected for a certain voice was summed and divided by the number of times it was presented (i.e., 18).

5.5.2 Results

5.5.2.1 By-Voice Text Responses

First, I report on responses for each AAC voice on the open-ended questions about demographics. The proportions of each response code per VOICE are presented in Table 5.8. In terms of guessing ethnic background, participants tended to respond with “white” or related labels (e.g., “Caucasian”, “European”): All voices were coded as White at least 20% of the time. At the same time, all voices received at least one response for Asian as well.

Four voices were primarily identified as White (Sp4, Sp2, Sp7, Sp11), having been labeled as such at least 75% of the time (or 14 out of 18 responses). Four voices were mainly split between Asian and White, although White was still the more common response (Sp3, Sp6, Sp8, Sp10). One voice was mostly described as White or Other (Sp5), including responses for Black (n=3) and Hispanic/Latin (n=2). Two voices in particular were more often coded as Asian than White (Sp1, Sp9), although Sp1 elicited more Other responses, being described as “Hispanic” (n=3) some of the time.

In terms of guessing where the speaker grew up, North America was the most common answer, both overall and per speaker. The majority of speakers were nearly always identified as being from the US or Canada (at least 89% of the time). One speaker tended to be secondarily identified as from Europe (Sp3) while another was often identified as growing up in Asia (Sp9). Finally, most responses describing the speaking style or voice did not mention accents or accentedness. However, one speaker was described with a Foreign Accent half of the time, mainly involving mention of a “slight accent” (n=4) or “Asian accent” (n=3).

Overall, even though all voices were in fact Asian American or Canadian, there was a general tendency to guess that the speaker was white. Speakers were most commonly interpreted as growing up in North America, which in this case was consistent with the general background of the sample. While each voice appeared to elicit somewhat different interpretations of ethnic background, a few voices appear to sound more likely Asian than the other voices, at least to some of the respondents.

Question	Response Code	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Sp7	Sp8	Sp9	Sp10	Sp11
Ethnic Background	Asian	0.44	0.06	0.39	0.06	0.11	0.28	0.06	0.33	0.72	0.33	0.11
	White	0.33	0.89	0.61	0.94	0.56	0.61	0.89	0.56	0.22	0.56	0.78
	American	0.06	0.00	0.00	0.00	0.00	0.06	0.00	0.06	0.00	0.00	0.00
	Other	0.17	0.06	0.00	0.00	0.33	0.06	0.06	0.06	0.06	0.11	0.11
Regional Origin	Asia	0.11	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.39	0.00	0.00
	North America	0.89	0.94	0.61	1.00	0.94	0.94	0.94	1.00	0.44	0.94	1.00
	Europe	0.00	0.06	0.28	0.00	0.00	0.06	0.06	0.00	0.06	0.06	0.00
	Other	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.11	0.00	0.00
Speech Style	Foreign Accent	0.06	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00
	Mainstream Accent	0.00	0.00	0.00	0.11	0.06	0.06	0.00	0.00	0.00	0.06	0.00
	Other	0.94	1.00	0.94	0.89	0.94	0.94	1.00	1.00	0.50	0.94	1.00

Table 5.8: Proportions of coded responses for open-ended questions about ethnic background, regional origin, and speech style, averaged across all responses for each of eleven AAC voices. For ease of visual interpretation, values above 0.2 are shaded in grey.

5.5.2.2 By-Voice Ratings

To assess more gradient social impressions of each voice, I turn to the results of the Likert and semantic differential scale ratings. The mean likelihood ratings for different demographic variables are displayed in Figure 5.6 while the mean social characteristic ratings for each speaker are displayed in Figure 5.7 (see Appendix B for a table of means per speaker).

On average, respondents rated the voices as neither likely nor unlikely to be ASIAN (M=3.70, SD=2.08) while being somewhat likely to be WHITE (M=5.22, SD=1.70); they also rated the voices as somewhat unlikely to be either LATIN (M=2.96, SD=1.72) or BLACK (M=2.52, SD=1.53). Ratings were relatively variable for each voice, however. In particular, compared to the open-end responses, we can see a more gradient distribution of voices being perceived as more likely Asian or white.

Only two voices were rated as overall likely to be ASIAN (Sp9: M=5.39; Sp1: M=4.89), as well as being rated as more likely to be ASIAN than WHITE (Sp9:

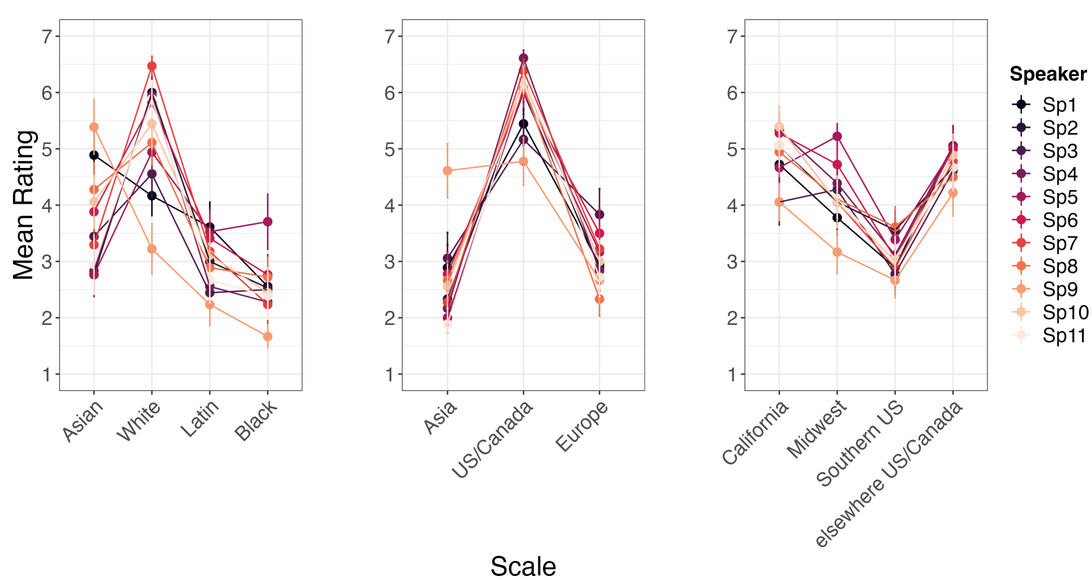

Figure 5.6: Parallel coordinates plot showing the mean values (error bars represent standard error) of the Likert scale ratings per VOICE.

M=3.22; Sp1: M=4.17). Sp9 appears to be consistently interpreted as Asian specifically. In comparison, the pattern for Sp1 is less lopsided, and this speaker was also rated as most likely to be LATIN out of all the speakers (M=3.61), suggesting a somewhat ambiguously non-white perception that listeners could not fully place. Both of these patterns match the general conclusions from the open-ended responses.

The other voices varied in their combination of race/ethncity ratings, and the voices least likely to be perceived as ASIAN were not necessarily the same voices most likely to be perceived as WHITE. For example, the two speakers rated the most likely to be ASIAN after the two discussed above (Sp8: M=4.28; Sp10: M=4.06) were also rated as somewhat likely to be WHITE (Sp8: M=5.11; Sp10: M=5.44), more so than a few other speakers.

In general, the six voices (Sp1, Sp3, Sp6, Sp8, Sp9, Sp10) that were somewhat consistently labeled with an ASIAN racioethnic label in the open-ended questions (at least 25% of the time) were also rated as relatively likely to be ASIAN in this task. Additionally, the voices (Sp2, Sp4, Sp7, Sp11) most consistently described as WHITE (at least 75% of the time) were also the most likely rated as WHITE in this task, demonstrating convergent results on the whole.

Average ratings for where speakers grew up were relatively consistent, indicating that speakers were considered likely to have grown up in the US/CANADA (M=5.90, SD=1.29) and somewhat unlikely to have grown up in ASIA (M=2.66, SD=1.68) or EUROPE (M=3.01, SD=1.65). The only speaker that was considered at least somewhat likely to have grown up in ASIA was Sp9 (M=4.61); they were also rated as similarly likely to have grown up in the US/CANADA (M=4.78). This pattern was again similar to those found for the open-ended responses. In terms of particular regions of the North America where the speaker was likely to have grown up, respondents generally interpreted the speakers as being from CALIFOR-

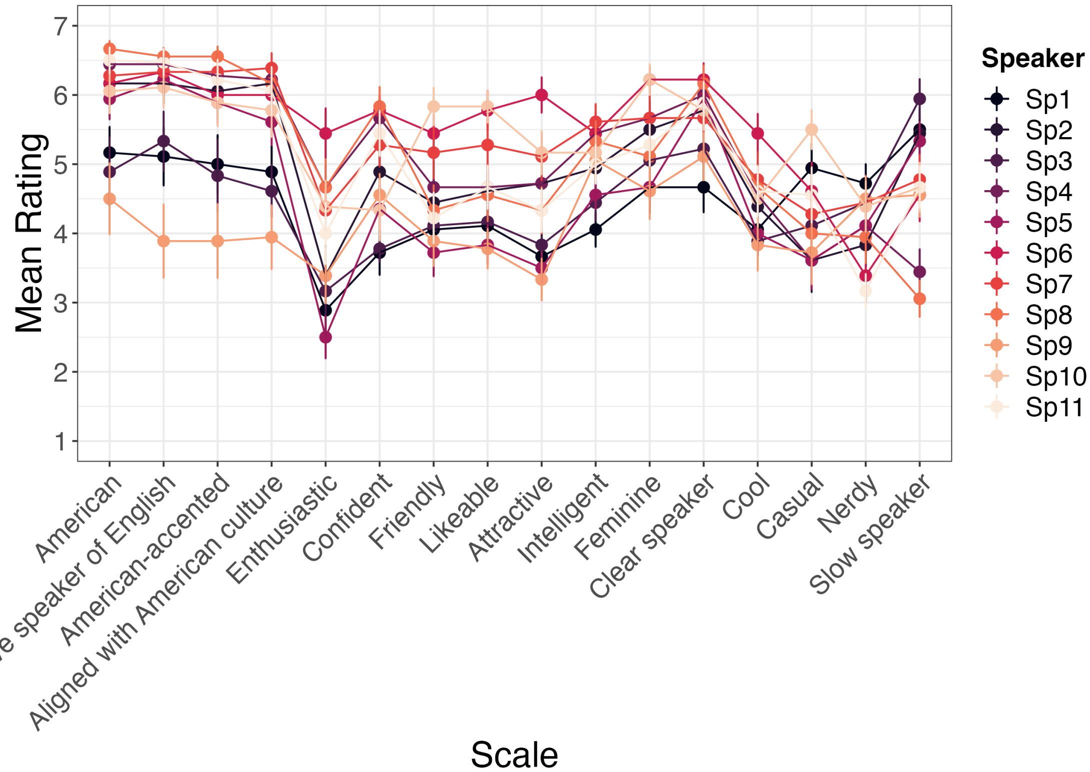

Figure 5.7: Parallel coordinates plot showing the mean values (error bars represent standard error) of the semantic differential scale ratings per VOICE.

NIA (M=4.90, SD=1.51) or ELSEWHERE IN THE US/CANADA (M=4.77, SD=1.60) compared to being from the MIDWEST (M=4.18, SD=1.55) or the SOUTHERN US (M=3.09, SD=1.52).

Moving on to the the semantic differential scales that measured various aspects of cultural alignment, personal traits, and stylistic expression, Figure 5.7 shows that while the ratings on Culture scales revealed a distinct pattern across voices, there was a wide range of responses on the various Trait and Style scales; here, I focus on the culture-related results. Overall, the majority of voices elicited high-to-very high ratings for AMERICAN (M=5.89, SD=1.54), NATIVE SPEAKER OF ENGLISH (M=5.90, SD=1.53), AMERICAN-ACCENTED (M=5.72, SD=1.71), and ALIGNED WITH AMERICAN CULTURE (M=5.62, SD=1.54). Thus, most of the speakers were seen as American, speaking English natively with an American accent, and aligned with the local culture. Two voices (Sp1, Sp3) were in a tier below, receiving moderately high average ratings for the four American-related scales. These speakers seem to be perceived as somewhat less American, linguistically and culturally. The final voice (Sp9) was rated on average close to the midpoint on the same four scales, indicating that they were interpreted as the least American of all the speakers.

5.5.2.3 By-Voice Selections

Lastly, I present the the photo selection results that more holistically assess how social impressions of the voices relate to the targeted personae represented by photos. Figure 5.8 visualizes the proportion of PERSONA selections for each VOICE as a heatmap. As multiple selections were allowed, the total proportion of selections for each voice may exceed a value of one. At a glance, each speaker elicited a different pattern of responses, suggesting that there were a variety of impressions across voices. All voices received at least one selection per PERSONA group, with the exception of Sp5 whose voice did not elicit any EA responses. Based on visual inspection of the data focusing on the top PERSONA group(s) selected per voice, three main patterns of responses representing broadly distinct social impressions can be discerned from Figure 5.8.

The first observation is that three voices (Sp1, Sp8, Sp9) had a clearly dominant Asian persona association, although the specific distribution of responses was variable. Of these, two speakers (Sp1, Sp8) were most associated with the MA persona such that respondents selected a MA photo option over half of the time; Sp1 also received IA selections half of the time. The third speaker's voice (Sp9) elicited a relatively high proportion of selections across all three Asian personae, but responses were slightly more concentrated on the EA and IA personae compared to the MA persona. The MA selections for Sp2 and Sp8 represent the most concentrated responses for any specific persona across all the voices.

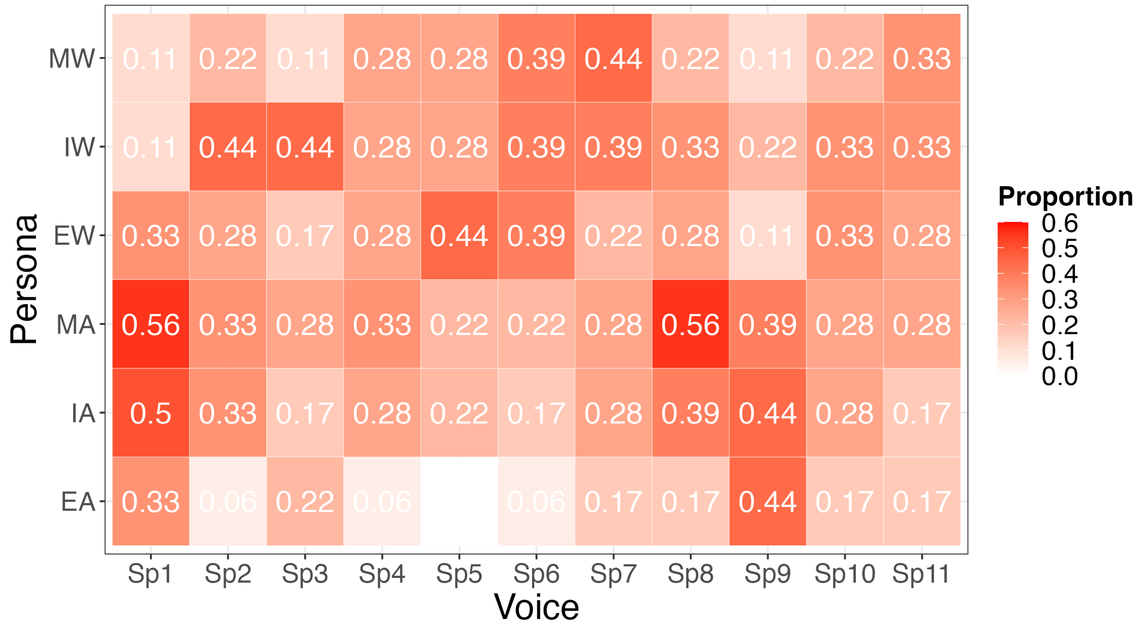

Figure 5.8: Heatmap showing the proprtion of times a photo from each PERSONA was selected per VOICE.

The second group of voices (Sp2, Sp3, Sp5, Sp6, Sp7) had a prominent association with one or multiple of the three white personae. Most of these speakers elicited photo selections that were concentrated on one persona: Sp7 was most associated with MW photos (and secondarily with IW photos), Sp2 and Sp3 were most associated with IW photos, and Sp5 was most associated with EW photos. In addition, one voice (Sp6) was most commonly associated with all three white personae, with selections evenly distributed across the three options.

The last general pattern we see involves the three voices (Sp4, Sp10, Sp11) that did not appear to have a dominantly associated persona. These speakers received a range of responses across both white and Asian personae photo options at lower concentrations.

At a high level, this data suggests that there were voices that sounded more Asian or more white, sometimes even sounding like a particular type of person within that racial/ethnic category. Moreover, there were voices that sounded more ambiguous to listeners in terms of racial/ethnic background. However, this overview represents just a simplified view of the general patterns, and the impressions of specific voices were often not clearly tied only to one racial/ethnic category on the whole. In particular, the voices often identified as white were also identified as Asian a sizable proportion of the time, which suggests that they still were considered potentially Asian American, even if there was an apparent bias for selecting white photos.

Further, these photo selection responses did not always line up with responses in the previous tasks. When compared to the demographic rating results, the three voices linked to Asian photos were indeed identified as the most likely to be Asian. However, they differed on the Culture-related ratings: While Sp1 and Sp9 were perceived as less American to varying degrees, Sp8 was rated along with the rest of the speakers as very much culturally American. Because Sp1 and Sp8 were instead the two that patterned similarly with regards to greater MA photo selections, the current picture of how voices may be connected to macrosocial and microsocial categories is unclear; whether these inconsistencies are due to task differences or underlying complexities in how voices are socially interpreted remains to be seen.

5.5.3 Interim Discussion

Overall, hearing different voices reading the same set of words led to different social impressions of each speaker. There were voices that were consistently identified as sounding Asian, as well as voices that were at times identified specifically as sounding like American-raised Asian individuals. Despite this, the association of voice to persona appeared not to be completely straightforward, at least within the constraints of this study.

One key finding from listeners' evaluation of demographic background was that speakers who were systematically identified as ethnically Asian were not necessarily also identified as being from Asia or foreign accented. In other words, when providing open-ended descriptions or likelihood ratings, at least some participants appeared to interpret a subset of voices as specifically belonging to Asians who grew up in North America and spoke with an American accent. This tracks with responses to the MA persona photos in Survey 1b, particularly for Sp1 who was more often identified as Asian than white while also being consistently identified as growing up in North America. The other voice that was most identified as Asian, Sp9, was considered to have grown up in either Asia or North America, which lines up with the response pattern found for the EA persona photos. This suggests that voices, or speech-based cues, can be associated with the same demographic categories that photos representing each targeted persona are.

Unlike demographic impressions, however, the results for ratings of social characteristics and selections of photo-based personae did not line up as neatly with the previous survey findings. For all voices, ratings related to being American-raised were closely associated with ratings about being American-aligned. This is unlike Survey 1b where MA photos were rated as overall more American-raised than they were American-aligned, which I interpreted as capturing an element of biculturalism. Thus, while voices did in fact show a three-way distinction in ratings for the American-aligned scale similar to that found for photos, they did not show the same pattern in ratings for American-raised scores.

Additionally, even though photos were grouped into persona categories based only on the culture-related scale ratings, voices that patterned similarly on those same rating scales were not necessarily linked to the corresponding persona photos. That is, while photo ratings directly map onto photo groups, voice ratings did not map onto photo groups as directly.

The most prominent example of an apparent misalignment between ratings and photo selections was seen in Sp8. This voice was rated as only slightly likely to be Asian in terms of absolute ratings (while being rated somewhat likely to be white). Of all 11 voices, they elicited the highest ratings on the three American-raised scales along with being third highest on American-aligned. These outcomes would seem to point towards an expected selection bias for white persona photos.[11] However, the photo selections were concentrated on the Mainstream Asian persona instead. This matched the rate of Mainstream Asian selections for Sp1, whose voice was strongly perceived as Asian North American and was rated as one of the least American, with moderate American-raised and American-aligned scores.

Looking to open-ended descriptions of the voices or speech style did not provide any relevant clues to why Sp1 and Sp8 in particular were most associated with a mainstream-oriented or bicultural Asian American persona. These two speakers were described quite differently (e.g., "slow, monotonous, nasal" vs. "fast, energetic, confident") and did not elicit any unique shared descriptors. To the extent that photo selections can be reliably interpreted as linking voices to specific personae, listeners appeared to draw on more subtle phonetic aspects of the speech samples to decide on which AAC persona was the best match.[12]

This example showcases a few insights from this survey. First, certain voices may be associated with a specific AAC persona, in this case a mainstream-oriented Asian persona compared to an ethnically-oriented Asian persona or a white persona. Second, ratings of how American-raised or American-aligned a

[11]Other photos rated in the top five on the culture scales, like Sp8, were either associated mainly with white persona photos (Sp2 and Sp7) or evenly associated with both white and Asian personae (Sp 4 and Sp11). Note that Sp8 did elicit more white photo selections than Sp1 and Sp9, the two who were mainly associated with Asian photos; this at least is consistent with this voice's higher ratings for white and American impressions.

[12]It is possible that specific voices were disproportionately tied to specific photos rather than to a persona category (e.g., due to close matches on non-target stylistic elements). This is difficult to judge in the current task as photos were randomly selected for presentation per participant, meaning photos varied in the number of times it happened to be paired with each voice. What can be said, at the very least, is that neither of these patterns was driven completely by one photo: Both speakers elicited at least two selections per MA photo. Regardless, that a voice may be associated with a particular subset of the persona category does not necessarily imply that the voice is not associated with a MA persona at all.

voice sounded did not always match with how likely they were associated with a mainstream Asian persona photo. Third, a voice may be more likely associated with an Asian identity in the presence of naturalistic photos than when considered in the abstract (e.g., race/ethnicity ratings).

That last point dovetails with post-task comments from two participants expressing the sentiment that it felt easier to link voices to an Asian (American) identity when there were concrete faces. One wrote that "The pictures upended my initial feelings [that the voices were white] and I could better imagine putting the voices to the faces shown, which was quite interesting." The other remarked that "once I saw faces I could more likely imagine the voices coming from [A]sian [A]mericans." These outcomes suggest that a reflection on the types of tasks involved in ethnic identification studies may be warranted, especially with an eye towards how results from those tasks may be generalizable to more realistic situations where stylistic information is available.

Coming back to the comparison of Survey 1b and Survey 2, why did ratings for the same American-related scales pattern differently for voices and photos? While providing judgments of voices compared to photos may simply be inherently different tasks, the differences in stimuli between the two surveys may have been a major factor. Unlike the photo stimuli, speakers in this survey were recruited as a convenience sample based on requirements for the experimental stimuli recordings and therefore were not selected specifically to vary in style. The speech samples were also designed to be as controlled as possible, using recordings of the same words based on read speech. This meant that the auditory stimuli were not naturalistic, further limiting potential cues to speech style and an individuals' social characteristics. In other words, while the survey instrument and overall methods were relatively well-matched, the stimuli were not particularly comparable. Moreover, the photos selected to represent each persona were very limited. Some of the variation in rating and photo selection results may be due to voices being associated with particular photos for other stylistic reasons that could be related to the ratings of personal traits and stylistic expression. To address these possibilities, it would be interesting to conduct the same study with a larger set of photos and voices along with more comparably designed speech samples that varied in style.

Before moving on, I briefly touch on the aspect of magnitude. Why did listeners seem to agree most on their selections of Mainstream Asian photos for the two voices that were considered to match with this persona? One possibility is that the respondents could themselves be considered mainstream-oriented Asian Americans and therefore had the strongest or most consistent expectations about what that might sound like. If so, this would support an interpretation where listeners do indeed have linguistic ideologies of what a mainstream-oriented Asian American voice sounds like, and was able to apply that knowledge linking speech style to visual style to pick photos associated with that persona.

Overall, although the noisy nature of the voice-to-face mappings do not provide conclusive evidence that specific types of voices are ideologically tied to specific AAC personae, the relative consistency of certain response patterns across participants hint that this could be the case; given the controlled nature of the stimuli, these patterns might be further enhanced if additional stylistic information was available in the auditory signal. Results show some degree of consensus on speech samples that were more likely associated with Asian speakers than white speakers, and specifically more associated with a mainstream-oriented AAC persona than an ethnically-oriented AAC persona.

5.6 General Discussion

The main goal of this study was to describe AAC personae that contrast along a continuum of acculturation, including their links to speech styles. After probing social evaluations of photos and grouping selected photos into emergent personae categories, I examined social evaluations of AAC voices and connected these back to the photo-based personae. The results generally demonstrate that Californian Asian Americans have distinct and detailed sociolinguistic associations for AACs who are perceived as more ethnically-oriented or more mainstream-oriented. In what follows, I discuss in more detail how these recognizable ideological constructs may be characterized and how these characterizations link to the broader literature surrounding ideologies about Asians in North America.

5.6.1 Operationalizing AAC Personae

The photo-based social evaluation tasks revealed that while judgments of race/ethnicity and regional origin were quite categorical, Asian-perceived photos were gradiently interpreted as more culturally Asian or culturally American. This outcome is consistent with the acculturation continuum as described by Pyke and Dang (2003) using qualitative interviews of second-generation Korean and Vietnamese Americans. The current results can be seen as quantitative evidence corroborating this ideological intra-ethnic relationship in a slightly more diversified Asian American sample that includes other ethnicities and immigrant generations. Referencing the social types labeled by Pyke and Dang (2003), the emergent ethnic- and mainstream-leaning AAC categories created from the photos in Survey 1b may conceptually align with a FOB Asian persona and Normal Asian persona; I discuss these categories in more detail below.

Compared to photo judgments, voice-based social evaluations were much more variable. This could be because the interpretation of stylistic information from voices compared to visuals tend to be less consistent across individuals, but in this case, it is likely that the relatively controlled nature of the speech samples played a large role in constraining stylistic interpretation within this survey specifically. Regardless, response patterns indicate that listeners' linguistic ideologies include ideas about what speech sounds like when spoken by Asian Americans who grew up in North America. Voices identified as such were distinct from those who tended to be identified as sounding like white Americans or like Asians who may have grown up in Asia. Recall that all speech samples were recorded by self-identified AACs. This result thus mirrors a general finding from perceptual ethnic identification studies showing that some but not all AAC speakers can be consistently identified as Asian (e.g., Hanna, 1997; Newman & Wu, 2011; P. Wong & Babel, 2017; A. Cheng & Cho, 2021). Whether identification responses are accurate or not is not of interest here. Rather, I see this body of literature as jointly indicating the existence of common ideologies about AAC ethnolinguistic variation which listeners may draw upon during speech processing. Even if there may not be a pan-ethnic "Asian American English" grounded in production (Reyes & Lo, 2009), there certainly seems to be a potentially pan-ethnic "Asian American accent" rooted in ideologies.

Whether speech alone can be linked to more narrowly-defined AAC personae, like FOB Asian or Normal Asian persona, is less clear from this data. Some speech samples did elicit converging responses for a portion of respondents, especially for a mainstream-oriented Asian persona based on photos. But, the fact that selections of visually-cued personae did not always correspond to the expected demographic and social attribute impressions for the same speaker (specifically for Sp8) further hints at complex relationships. This is not too surprising given the lack of stylistic variation available in the audio, among other limitations of the stimuli and differences between tasks; as such, I refrain from drawing any conclusions on this topic based on these data. Future research using naturalistic speech samples, like excerpts from interviews (e.g., Campbell-Kibler, 2007; Drager et al., 2021), could be better poised to explore more thoroughly what collection of linguistic elements may be linked to differently-acculturated AAC personae beyond general associations of foreign-accentedness with FOB Asians. Another approach would be to test whether the presence and absence of specific linguistic features elicit different AAC personae responses—this is the approach I take in the following chapter.

By providing details about the socioindexical fields of each persona, these results help to flesh out ideological constructs of differently-acculturated AACs for present-day Californian Asian Americans. Starting with visual cues, the more ethnically-oriented persona appeared to be associated with Asian makeup trends (e.g., straight brows; pink- or red-toned lip color; light, natural eye makeup) and sleek, styled hair. Clothing may be considered more stylish, feminine, or associated with Asian fashion trends (e.g., blouses, high neck tops). Based on social attribute ratings, these photo subjects were distinct for being perceived as relatively non-American, non-native in English, foreign-accented, and unclear in speech. In addition, they were seen as more intelligent, nerdy, and formal; the last of these seems to reflect the relatively polished or put-together styling of makeup, hair, and clothing in these photos. Overall, these observations align with the idea of the FOB Asian being someone who is more closely affiliated with Asian culture, as well as someone who represents negative Asian stereotypes (e.g., "nerdy" or "model minority" Asian Americans). Note that Pyke and Dang (2003) described the ideological "FOB" category as internally inconsistent and "the repository of all negative notions of Asian Americans perpetuated by the dominant society" (159). If this holds true, the social type captured in this study would represent only one possible instantiation of a FOB Asian persona.

A comparatively mainstream-oriented persona was attributed to photo subjects wearing either little to no makeup or American-style makeup (e.g., arched brows; nude or natural lip color; thick, dark lashes). Their hair tended to look less intentionally styled (e.g., simple ponytails or buns) and were sometimes dyed with highlights or unnatural bright colors. Clothing often included casual wear or athletic apparel (e.g., tank tops, sweatshirts and hoodies), though not all did. While there seemed to be multiple sub-styles represented, all of these photos could be characterized under the umbrella of Asians who adhere to some form of American cultural norms in appearance or implied behavior (e.g., outdoor or athletic activities).

Compared to the previous persona, social attribute ratings mainly identified these individuals as being relatively American, as well as speaking English natively and clearly with an American accent. They were also considered to be stylistically more casual and less feminine, likely driven by the subset of photos with particularly minimal makeup, unstyled hair, and relaxed attire. Although some trends could be extracted from the group as a whole, the varied styles associated with this bicultural category is consistent with Pyke and Dang's (2003) account of the "normals" being a catchall group without stable, independent traits; rather, most second-generation Asian Americans would self-identify with being "normal" in order to reject association with "FOB" or "whitewashed" labels. Viewed this way, there may not be any single Normal Asian persona. However, even if vague and heterogeneous, the bicultural identity still appears to exist as an ideological construct in some form, so I employ the Normal Asian persona as an imperfect category to capture this idea.

Like "FOBs" and "normals", the Whitewashed Asian persona was identified as a recognizable AAC social type in both Pyke and Dang (2003) and Chapter IV. However, it did not emerge as a distinct type of AAC based on photo evaluations. While the major distinguishing features of "whitewashed" individuals seem to be behavioral (e.g., "acting white", being unfamiliar with ethnic practices, socializing with and dating non-Asians) and linguistic (e.g., not speaking heritage languages, using white-associated speech styles), Pyke and Dang's (2003) informants also mention appearance-related stylistic elements (e.g., tanned skin, dyed blonde hair, casual or American-coded fashion such as baseball caps and footwear like Doc Martens). One possible explanation for the current outcome is that there are fewer distinctive visual features indexing this social type in the present day, thus no photos were able to be specifically identified as a Whitewashed Asian persona. Alternatively, the sample of photos in this study simply did not include enough photos of this type to find a consistent pattern, compounded by the restricted view of portrait-style photos that limit the available stylistic information (e.g., other clothing items).

In general, no one visual element was necessarily found in all photos associated with a persona, even within the narrowed photo sets used in Survey 2. Rather, holistic impressions seem to be created by selected visual features, such as makeup and hair, or makeup and clothing.

5.6.2 AAC Personae and Racialized Ideologies

Stepping back from the details of each persona, the results viewed at a higher level connect to various findings related to ideologies of Asians and their speech in a North American context. I found that photos of Asian-perceived individuals elicited relatively consistent judgments about their social and linguistic attributes depending on visual styles. Thus, a person's appearance was ideologically tied to what kind of person they were and how they were expected to speak, by virtue of both facets being associated with a particular recognizable type of person—a persona.

D'Onofrio (2019) detailed similar outcomes based on social evaluations of three photos. There were a number of methodological dissimilarities, including use of differently-gendered stimuli and a different participant pool: D'Onofrio's (2019) stimuli involved individuals perceived as men and the sample was mainly composed of white Americans located anywhere in the US. Despite these differences, our two sets of results converged on the following: Asian faces perceived as less American tended to be interpreted as more foreign-accented than Asian faces perceived as more American, and the former were also seen as more nerdy than white faces perceived as more American. The current study additionally assessed and found that Asian faces perceived as less American were likely perceived as unclear and non-native speakers of English as well.

These findings do not stand alone, but are enmeshed in more generalized stereotypes of Asian-racialized individuals. AACs are commonly linked in stereotypical portrayals to foreignness and non-Americanness represented by "forever foreigner" ideologies, but at the same time, "honorary white" ideologies suggest that AACs are highly assimilated and have mainstream accents (Tuan, 1998; Reyes & Lo, 2009, see also Devos & Banaji, 2005). Other Asian-racialized social attribute stereotypes include purportedly positive traits, like being intelligent, hard-working, concerned for others, and honest, as well as negative traits, like being poor communicators, unconfident, unassertive, and lacking leadership (e.g,. Lindemann, 2003; Hosoda et al., 2007). The "model minority" stereotype, in particular, emphasizes intelligence as a characteristic embodied by Asians (e.g., Bauman, 2013). Such stereotypes are complexly intertwined with attitudes about Asian-accented speech and appear to also influence the current social evaluations results.

Notably, previously-reported judgments of individuals being poorer speakers or communicators based on hearing Asian-accented speech (Lindemann, 2003; Hosoda et al., 2007) may be related to impressions in the current study about individuals being less clear speakers. Not only did respondents apply this attribute to culturally Asian-looking photos, they also showed this pattern when hearing speech samples, such that the three voices rated least likely to be American-accented or native speakers of English were the same three rated as the least clear speakers. The shared patterning of Americanness, American-accentedness, native speakerhood, and clarity of speech seem to comprise a cluster of (ethnically-oriented or foreign) Asian speech ideologies that can be activated both visually and auditorily. A similar collection of attributes was found to pattern together when photos and speech were presented together in D'Onofrio (2019).[13]

Beyond negative judgments of speaking ability, Asian-accented speech has been found to elicit more negative ratings than mainstream accents on many attitudinal scales. Similar to the findings of Bauman (2013), the three speakers heard as most foreign-accented in the current study were rated less positively on (nearly) every trait in comparison to the average. This pattern encompassed lower ratings on dimensions of status (e.g., intelligent), solidarity (e.g., friendly, likable), and dynamism (e.g., confident, enthusiastic), as well as being rated as less feminine, more nerdy, and slower speakers. While previous studies varied as to the specific attributes that were assessed, as well as the different accents (e.g., Korean, Vietnamese) and listener populations (e.g., non-Asian Americans, ethnically diverse Americans) involved, those relating to the dimension of status (e.g., ambitious, successful, intelligent) have been reported as consistently downgraded for Asian accents (Lindemann, 2003; Hosoda et al., 2007; Bauman, 2013). Insofar as the attribute of intelligence alone can represent status, the current results find the same effect, and as pointed out by previous scholars (e.g., Bauman, 2013), the impression of lower intelligence runs contrary to typical Asian-racialized stereotypes.

[13]Interestingly, a seemingly contradictory effect—voices evaluated as *clearer* when paired with a less American-perceived Asian photo—was found in D'Onofrio (2019). One speculative explanation was that in both viewing an image and listening to speech, the interaction of expecting unclear speech based on a photo before hearing clearer-than-expected speech may have led to contrastively higher subjective ratings of clarity. This explanation is somewhat similar to Y. Zheng and Samuel's (2017) proposed context effect on perceived accentedness that found that the presence of strongly accented speech paired with an Asian face led to reduced accentedness ratings of ambiguous speech paired with an Asian face.

On the whole, while this study was not specifically designed to assess attitudes towards Asian-accented speech, the current results appear to align with the body of literature that finds generally negative associations attributed to Asian accents. It suggests that these attitudes appear maintained even in a fully Californian Asian American sample in the present day.

However, one potentially interesting aspect of the data is the differences found between social evaluations when *seeing* Asian-racialized faces considered foreign-accented and when *hearing* AAC voices considered foreign-accented. Unlike voices, ethnically-oriented Asian faces were rated quite positively on the various personal traits even though they were interpreted as more foreign-accented, less clear speakers, and more nerdy; further, photos read as more culturally Asian *were* accorded higher intelligence, in alignment with stereotypes like "model minority". Mainstream-oriented Asian faces were rated similarly while not being associated with foreign-accentedness.

It must be acknowledged that one-to-one comparisons between the current study's photo- and voice-based surveys may be ill-advised as stimuli were rather dissimilar in quality.[14] For example, faces were all smiling and generally attractive (as might be expected of stock photo models), while voices were more serious and generally sounded unenthusiastic (as might be expected in the context of reading words for an experiment). Still, the fact that faces could be associated with both a presumed Asian accent and positive characteristics while actual Asian accents are negatively evaluated may speak to complexities in understanding representations of and access to stereotypes based on visual or auditory information.

The current results may also help to contextualize reported effects of Asian faces inducing impressions of greater foreign accentedness when hearing speech.

[14] I would also note that the specific accent of each speaker was not necessarily identified as Asian by all listeners, based on different likelihood ratings for other racioethnic identities. This may or may not be related to differences in ratings across voices. In contrast, all the faces were definitively identified as Asian-perceived and the foreign accent was presumably assumed to be Asian.

Specifically, studies have found that (East) Asian faces presented alongside speech (usually mainstream North American speech labeled as L1 or native American English) resulted in higher subjective ratings of accentedness compared to seeing a white face, whether viewing static images or dynamic videos (Rubin, 1992; Yi et al., 2013; M. Babel & Russell, 2015; Y. Zheng & Samuel, 2017). The actual effect seems to vary in nature, such as whether seeing an Asian face actively increases accentedness ratings, whether seeing a white face reduces accentedness ratings (M. Babel & Russell, 2015), or both at the same time (Yi et al., 2013). Moreover, the effect is malleable, evidenced by variable results based on demand characteristics (e.g., seeing photos vs. video) and the presence of contextual speech affecting strength and direction of judgments (Y. Zheng & Samuel, 2017). This suggests that differences in visual styles depicted by the images could also modulate results, potentially in complex ways. The inclusion of stylistic variation in the images may be related to why, unlike the studies cited above, D'Onofrio (2019) did not find a difference in accentedness ratings when pairing Asian faces with voices, even for the photo that was read as more foreign-accented.

Recent studies tend to posit an explanation where the macrosocial racialized category of "Asian" is stereotyped as foreign; thus, seeing an "Asian" face results in expectations of foreign-accentedness which may further affect comprehension if there is an expectation mismatch with incoming speech (e.g., M. Babel & Russell, 2015; McGowan, 2015; M. Babel & Mellesmoen, 2019). However, as the current study's social evaluations demonstrate, different expectations may be foregrounded depending on ideologies connected to how the person looks—whether these ideologies are, on an individual level, sourced from interactive exposure to or cultural metacommentary about different types of AACs who have assimilated to American culture to differing extents. Essentially, rather than always expecting a foreign accent based on generalized stereotypes—or rather than being equally likely to accept either a foreign or mainstream accent given experience with both (e.g., M. Babel, 2022)—we should expect different Asian-racialized individuals to be systematically interpreted as likely or not likely to carry a foreign accent. Beyond the dimension of acculturation, particular visual cues could also more strongly evoke other social styles that may come with its own set of linguistic expectations. Even the use of photos intended to control for or limit style-related elements can never be completely devoid of social interpretation (for more discussion, see D'Onofrio, 2019), and selecting a photo of purportedly ambiguous social qualities could potentially result in unpredictable or divergent expectations by different participants (e.g., some may expect foreign-accentedness while others do not).

Portrayals relying on the assumption that any Asian individual will be associated with the same social and linguistic attributes are overly simplistic and frames AACs as a monolith. One major issue is that this may not characterize ideologies, representations, and processing at an individual level. We can also consider this a matter of ecological validity: How well do conclusions about "Asians" as a macrosocial category generalize to realistic social interactions with AACs that are likely rich with stylistic and contextual information? The present results support D'Onofrio's (2019) call to reconsider approaches focused only on macrosocial category membership that may erase meaningful and systematic heterogeneity and perpetuate monolithic views of racialized groups.

5.6.3 Conclusion

Three social evaluation surveys in the context of AAC faces and voices reveal social personae differing along a continuum of acculturation. This included a more ethnically-oriented (culturally Asian) Asian American character, linked to FOB Asian and "forever foreigner" ideologies, and a more mainstream-oriented (bicultural) Asian American character, linked to Normal Asian and “honorary white” ideologies (Pyke & Dang, 2003; D’Onofrio, 2019). As personae, these ideological constructs are complexly but recognizably tied to sets of social attributes that include demographics, appearance, behavior, speech, personal traits, and more (e.g., Eckert, 2008a; D’Onofrio, 2020; Drager et al., 2021). In the next chapter, I use these operationalized personae in an experiment to test whether two specific phonetic features are associated with AAC speech, and specifically whether they may be associated with one AAC persona specifically.

CHAPTER VI

Ideology in Action: Personae, Awareness, and AAC Sociophonetic Variables in Social Categorization

6.1 Introduction

In Chapter IV, exploratory survey results hinted at /oʊ/-backing and /ð/-stopping being ideologically associated with AAC speech to some degree of awareness, as well as at the existence of various AAC ideological types in present-day Asian American consciousness. In Chapter V, I confirmed through social evaluation tasks that Asian Americans from California have access to detailed ideologies about social types of AACs linked to particular visual and linguistic styles, including those who are considered more aligned with Asian culture or American culture. Altogether, I have established that Asian Americans have linguistic ideologies about AAC speech, including beliefs about different types of Asian Americans who vary along a continuum of acculturation.

This groundwork sets up the /oʊ/-backing and /ð/-stopping variables in the AAC context as an opportunity to expand on previous sociophonetic research from two perspectives: (a) considering how ideological categories like personae may mediate sociophonetic representations and processing, and (b) testing how awareness mediates the encoding of sociophonetic associations and later perceptual behavior. The present chapter both directly assesses /oʊ/-backing and /ð/-stopping as AAC sociophonetic variables and investigates how the constructs of personae and awareness may be involved in sociolinguistic representations through a social categorization task.

6.2 Background

As reviewed in Section 2.1.2.1, various approaches to sociophonetic representations theorize that the way linguistic experiences are interpreted and stored in memory is mediated by listener attention, guided by ideology (e.g., G. J. Docherty & Foulkes, 2014; Sumner et al., 2014; Drager & Kirtley, 2016; D'Onofrio, 2021b). This helps to explain the body of literature that finds that social speech perception may be influenced by ideological elements like stereotypes (e.g., Niedzielski, 1999; McGowan & Babel, 2019) and personae (e.g., D'Onofrio, 2018b; Leigh, 2021), rather than simply being influenced by experience with direct, interactional speech (see review in Section 2.1.1). In addition, it aligns with literature positing that sociolinguistic salience or awareness (tied to attention) influences the presence or magnitude of sociophonetic perceptual effects (e.g., Drager, 2011; Alderton, 2020).

However, the proposed relationship between sociophonetic representations, awareness, and ideology has not been tested in depth. For example, while awareness has been predicted to increase strength of sociophonetic associations via salience and attention (Drager & Kirtley, 2016), the few studies exploring this connection use indirect or imprecise measures and find inconsistent results (see review in Section 2.1.3.2). Plus, although the social information that listeners map to phonetic detail may be better captured at individual levels by microsocial ideological constructs (personae) rather than macrosocial categories (D'Onofrio, 2020), relatively little work has explored this link in perception (see review in Section 2.1.3.1).

Separately, as discussed in Section 4.2.2, not only do we know relatively little about perceptual cues to AAC ethnic identification, we know even less about the extent to which specific cues are tied to awareness and whether cues are tied to specific personae but not others. While /oʊ/-backing and /ð/-stopping come up in metalinguistic commentary about (A. Cheng & Cho, 2021) and performances of (Chun, 2004) AAC speech, details about how they may factor into sociophonetic representations and perceptual processing of AAC speech is murky. To the best of my knowledge, no studies have systematically tested the effect of specific linguistic features on perceptual identification of AAC speakers.

6.2.1 The Current Study

Using the context of AAC speech, I examine two phonetic features, /oʊ/-backing and /ð/-stopping, that may be ideologically linked to AAC identities with different degrees of awareness (Chapter IV). I use targeted gradient measures of awareness to test predictions systematically at both group and individual levels. Based on acculturation being a relevant social dimension for AACs (Chapter V), I examine the association of these phonetic variables with Asian American personae judged to be more mainstream-oriented or more ethnic-oriented. This provides a novel context for testing the role of persona in sociophonetic perception. Simultaneously, this approach helps to establish a clearer and fuller picture of /oʊ/-backing and /ð/-stopping as sociophonetic variables indexing AAC identity. I bridge the gaps in our understanding by examining these two variables in depth, including testing whether their presence or absence influences AAC identification.

Working backwards, then, the first set of goals for this chapter target how /oʊ/-backing and /ð/-stopping factor into ideologies about AAC speech. One aim is to establish with more certainty whether /oʊ/-backing and /ð/-stopping are ideologically associated with AAC identity, and in particular, to what degree of awareness. Do people believe that these pronunciation variants are used by Asian Americans in an "Asian American accent" and if so, how accessible are these ideas or how prominent do they feel? Another aim is to test through behavior whether /oʊ/-backing and /ð/-stopping can be used as perceptual cues to AAC speech, which also indirectly assesses the socioindexical link between these variants and AAC identity. In other words, can hearing /oʊ/-backing and /ð/-stopping in a speech sample bias listeners to interpret the speaker as Asian American?

In doing so, I specifically explore whether /oʊ/-backing and /ð/-stopping are differentially linked to narrower, microsocial AAC subcategories within the macrosocial "Asian" category, focusing on personae that vary along a continuum of mainstream to ethnic cultural orientation. That is, I ask whether hearing /oʊ/-backing or /ð/-stopping can bias listeners to interpret the speaker as a particular type of Asian American, such as a FOB Asian or Normal Asian. By systematically testing the presence or absence of two phonetic variants on speaker identification behavior, this study will provide novel empirical evidence identifying specific cues that give rise to perceptual impressions of "sounding Asian".

The next set of goals test how ideology factors into sociophonetic representations and perception. The first of these aims to contribute to research on how microsocial categories (as opposed to macrosocial categories) may be linked to phonetic detail and mediate perceptual processing, testing the role of personae in sociophonetic perception in a novel context. Specifically, I ask whether personae-based information about AACs can influence sociophonetic perceptual behavior. The second aim is to test whether awareness (via salience and attention) strengthens the encoding of sociophonetic representations, which may further bolster activation of these associations during perceptual processing to influence behavior. In this case, I ask whether degree of awareness for AAC sociophonetic variables is correlated with stronger biases in perceptual behavior.

To address these various aims, I conduct a speaker identification task using a multi-speaker matched guise technique. Listeners will be asked to select who they are hearing from two photo options while the words they hear vary in pronunciation and the photos they see vary in persona. Each voice produces the same words with two pronunciations, and photos vary in visual style to represent different personae. In this way, we can measure whether there is a bias for selecting certain AAC personae when hearing /oʊ/-backing and /ð/-stopping. In addition, awareness of /oʊ/-backing and /ð/-stopping ideologies will be assessed via targeted gradient self-report measures on a questionnaire. By combining these two sources of data, we can assess whether awareness is related to differences in speaker selections. As in the previous studies, this study examines Californian AAC listeners; the possibility of high exposure to AAC speech in California was expected to provide ample variation in awareness of the features under study. The specific research questions are as follows:

1. Are /oʊ/-backing and /ð/-stopping ideologically associated with AAC speech and to what extent of awareness?
2. Does the presence of an AAC-associated phonetic variant bias listeners to identify the talker as having an Asian and/or ethnically-oriented Asian identity?
3. Is greater awareness of a feature's social association linked to larger speaker identification biases?

For the question about awareness, I ask both whether the effects of hearing AAC-associated phonetic variants on speaker identification is stronger for (a) variants with greater community-level awareness and (b) individuals with greater awareness.

6.3 Methods

This study employed a social categorization task with multiple speakers in order to assess whether listeners would be more likely to guess that a speaker is Asian, and in particular an ethnically-oriented Asian American, if they heard them produce backed /oʊ/ or stopped /ð/. The design was specifically an auditory two-alternative forced choice speaker identification task where the response measure was photo selections. Auditory stimuli provided the phonetic information of /oʊ/-backing and /ð/-stopping in the form of word recordings. Crucially, each word was produced by AAC speakers with two phonetic variants, representing either Californian or AAC pronunciations: Mainstream (fronted /oʊ/, fricated /ð/) and Non-mainstream (backed /oʊ/, stopped /ð/). Visual stimuli provided the social information of different AAC personae, represented by photos of people with different visual styles (e.g., clothes, makeup). These included three target personae: Mainstream Asian, Ethnic Asian, and Mainstream White. After hearing a word recording, participants were to select one persona photo out of two options. A greater proportion of selections for a particular persona based on hearing a particular variable would be taken as evidence of a sociophonetic association. Degree of awareness of the variables were collected separately in a questionnaire. In the following, a provide more detailed explanations of each aspect of the methodology.

6.3.1 Linguistic-Auditory Stimuli

Linguistic stimuli consisted of auditory recordings of isolated words extracted from read productions. I first provide an overview of the stimulus design. These stimuli involved eight linguistic variables each represented by 10 words, totaling 80 unique words (for full word list, see Appendix C). Of these eight variables, two were target variables, two were comparison (or distractor) variables, and four were filler variables. Words were recorded by 11 speakers, of whom six were target speakers and five were filler speakers.[1] Stimuli for target speakers included all variables while stimuli for filler speakers involved only filler words. The critical stimuli, then, were the subset of words containing the target variables (/oʊ/-backing and /ð/-stopping) spoken by target speakers.

One purpose of the filler stimuli in the experiment was to distract participants from the experiment's focus on the critical variables and AAC-related phonetic variation by presenting a larger variety of words and pronunciation variants. The other purpose was to incorporate a larger variety of voices into the task, including some that were expected based on norming to be most perceptually associated with varied race/ethnicity and persona impressions. In addition to distracting from the experiment's focus on specific pronunciations and personae by increasing speaker-related variation, the greater number of voices aimed to reduce participants' ability to recognize and track specific voices. This was designed to help avoid a potential strategy in the experimental task of selecting photos only based on voice recognition (e.g., selecting a photo based only on having selected it for that particular voice on an earlier trial) rather than a holistic judgment.

The two target variables were associated with AAC speech (/oʊ/-backing and /ð/-stopping). On the other hand, the two variables intended as comparison variables were potentially linked to white speaker identity as opposed to AAC identity. Specifically, these were the PIN-PEN merger and FEEL-FILL merger, two forms of linguistic variation documented in some varieties of Southern US English (e.g., Labov et al., 2006; Koops et al., 2008; Wolfram & Schilling, 2015; J. A. Jones & Renwick, 2021). The remaining four variables (including three vowels and one consonant to match the structure of the target and comparison variables) were designed as fillers. These were expected to likely be familiar to listeners from California, but

[1] Stimuli for 10 speakers are available from the author on request; one did not consent to their recordings being shared.

were not expected to be strongly associated with a particular race/ethnicity. The three vowel variables were linked to Californian English or the California Vowel shift, specifically /u/-fronting, /æ/-backing, and /æN/-raising (e.g., Wolfram & Schilling, 2015; A. Cheng et al., 2016). The consonant variable was /t/-deletion (i.e., cluster reduction), a non-regional variation that may be linked to a variety of meanings (e.g., Eckert, 2008a).

In more detail, to target /oʊ/-backing and /ð/-stopping variables, the critical stimuli were words containing /oʊ/ before voiced obstruents and words containing intervocalic /ð/. Comparison words contained pre-nasal /ɛ/ and pre-lateral /ɪ/ which are the phonological contexts for the PIN-PEN merger and FEEL-FILL merger. Filler words included /u/ before a coronal voiced obstruent or flap (/u/-fronting), pre-nasal /æ/ (/æN/-raising), /ae/ before a coronal voiced obstruent or flap (/æ/-backing), and a coda /st/ or /ft/ cluster preceding a consonant.

All words were bisyllabic or trisyllabic with initial stress. The target sound was either the vowel in the first syllable, or the medial consonant(s) immediately following the first syllable. Words containing /oʊ/ began with a CVC structure where the vowel was /oʊ/ and the following consonant was a (non-velar) voiced obstruent (e.g., *goaded*). Words containing /ð/ began with a (C)VCV structure where the intervocalic consonant was /ð/ (e.g., *featherweight*). Comparison and filler words were similarly constructed based on the specific phonological contexts involved. Ten words were selected per variable, resulting in 80 total words (Appendix C).[2]

Auditory stimuli were self-recorded by 11 linguists with phonetic training who self-identified as Asian American/Canadian women (see Section 5.5.1.1 for more

[2]Stimuli were simultaneously designed to be used in a word identification experiment where each word had a corresponding competitor word that began with the same or similar sounds (e.g., *federal* as the competitor for *featherweight*). Thus, word choice for this experiment was constrained to options that had a viable competitor word with comparable word frequency. Due to the narrow set of word pair options for /ð/ and /oʊ/, the maximum number of words per variable was limited to ten.

details). Two primarily described themselves as Asian American, while the remaining nine described themselves with an East Asian ethnicity[3]. They spent the majority of their time before age 18 in the US (n = 4), Canada (n = 5), or Asia (n = 2). Ages ranged between 23 and 42 (M = 29.7) at the time of recording. Speakers recorded themselves in a sound-attenuated booth (n = 5) or quiet room (n = 6). Recording equipment varied by speaker: Six recorded using external microphones into Audacity, four recorded using Macbook laptop microphones into Audacity, and one recorded using a Zoom H4n Pro Handy Recorder. Although remote self-recording resulted in recordings with varied audio quality, this approach allowed for the inclusion of a range of speaker voices which otherwise may not have been accessible given the specificity of the speaker inclusion criteria. Because the recordings were to be used as one group of perceptual stimuli rather than analyzed acoustically and/or compared to each other, this approach was judged as acceptable for the current purposes.

Prior to recording, speakers were briefed on the various pronunciations they were being asked to produce, followed by a practice session with the experimenter. The word list was structured in sets of three tokens, one set per target word (e.g., *featherweight*, *federal*, *featherweight*). The first token was always a target word read with a mainstream pronunciation (e.g., fronted /oʊ/, fricated /ð/). The third token was the same target word read with a non-mainstream pronunciation (e.g., backed /oʊ/, stopped /ð/). In between these two, the second token was a word that began with the same onset and had the same or a similar sound to the non-mainstream variant (e.g., pre-lateral /oʊ/, [ɾ]). This ordering was intended to help participants achieve the non-mainstream variant pronunciation in the target word more naturally, having just read a word with similar pronunciation. Speakers were

[3]I intended to include speakers of Southeast Asian background as well. However, although additional speakers were recruited for recording, the final subset of speakers selected for inclusion based on quality of recordings happened to not include anyone who identified as Southeast Asian.

instructed to avoid list intonation and produce each word as similarly as possible in terms of suprasegmental features. They were also asked to repeat each triplet at least three times.

Due to the variance in recording environments, recordings first underwent noise reduction using Audacity (Version 3.2.4) to minimize differences. Then, word boundaries were automatically identified via silent interval detection using Praat (Version 6.2.14, Boersma & Weenick, 2022). All recordings were manually reviewed by the author. Based on auditory judgment and visual inspection of spectrograms, one token per target word and pronunciation variant was selected for each speaker where possible. In general, a token was considered for selection if it was judged as a successful realization of the target pronunciation. However, if a token was auditorily judged as sounding particularly unnatural or forced, it was excluded. In addition, tokens were excluded if they included non-speech noises. Then, mainstream and non-mainstream pronunciation variants of the same word were judged together. For each target word, a pair of tokens were selected: These were judged to be different on the target variant but as similar as possible on non-target segmental and suprasegmental details (e.g., pitch contour, voice quality, perceived naturalness).[4] All selected tokens were then extracted and average intensity was normalized to 60 dB.

To confirm that the stimuli across variant conditions clearly differed on the target phonetic dimensions and only those, tokens were acoustically profiled in a number of ways. Word-level pitch and duration were extracted using ProsodyPro (Xu, 2013). Pitch tracking errors, such as extreme fluctuations in pitch estimates, were manually corrected where needed. Segmental boundaries were identified via

[4]In some cases, the same-word token pairs were very similar phonetically, while in other cases, there were noticeable differences. However, as the goal was to examine differences at the level of the linguistic variables and phonetic variants, I focused on ensuring that the average phonetic profile of, for example, the ten tokens of /ð/ words with mainstream variants was comparable to the ten tokens /ð/ words with non-mainstream variants. This was accounted for by profiling the acoustic properties for every token and aggregating by condition.

forced alignment using the Montreal Forced aligner (McAuliffe et al., 2017). Then, formants and durations for vowels were extracted using FastTrack (Barreda, 2021). Manual correction of segment boundaries and format tracks were also conducted as needed. Systematic auditory coding was used to assess the consonant variants. Two trained phoneticians (one of which was the author) rated each token on how clearly they represented one variant or the other using a 5-point scale (e.g., clearly fricated, weakly fricated, ambiguously fricated/flapped, weakly flapped, clearly flapped). Tokens were randomly presented and blocked by variable while coders were blinded to speaker, word, and intended variant. Overall, these measurements provide confirmation that the tokens selected for mainstream and non-mainstream variant conditions differ as intended: Non-mainstream variants of /oʊ/ are produced with more backed articulation, indicated by lower F2 (Table 6.1), while non-mainstream variants of /ð/ sound more like flaps than fricatives, indicated by higher ratings (Table 6.2).

During the token selection process, six speakers were designated as target speakers (listed in Table 6.3). These speakers all grew up in the US or Canada and were perceived as such; in terms of racioethnic impressions, these voices were variably interpreted as more likely white, more likely Asian, or relatively ambiguous, and included the two speakers most associated with a Mainstream Asian persona (see full social evaluation norming results in Section 5.5.2 and Appendix B).[5] Target speaker tokens for all eight variables (target and filler) were included as stimuli. All ten words were included per variable and per speaker with two exceptions where only nine words was included for a particular

[5]Sp1 re-recorded the full word list for this experiment after the auditory norming study reported in Section 5.5. This was done in order to better match with the other selected target voices on speech rate and energy, as Sp1's voice was rated as particularly slow and unenthusiastic relative to other target voices. As such, the word recordings used in the norming study for Sp1 differed from those used in the current study, and therefore the social characteristic ratings reported in Section 5.5.2 may not directly reflect the impressions of Sp1's voice in the current study. However, I expect that the overall impressions of this speaker are likely maintained as the voice characteristics did not change.

Word	Mainstream Variant (fronted)			Non-mainstream Variant (backed)		
	F1	F2	F3	F1	F2	F3
coding	579	1484	2680	523	1038	2834
goaded	570	1517	2603	541	1090	2725
hoses	568	1446	2716	521	1010	2885
modem	592	1424	2902	586	977	3005
podium	596	1444	2726	525	962	2904
posers	577	1396	2711	522	936	2924
robot	565	1402	2512	551	1030	2611
soberness	554	1586	2631	541	1108	2778
soda	612	1610	2631	602	1176	2701
sojourn	564	1550	2676	518	1061	2770
Grand Mean	578	1486	2679	543	1039	2814

Table 6.1: /oʊ/ auditory stimuli phonetic profile. Mean formant measurements at /oʊ/ nucleus (30% into the vowel) for Mainstream and Non-mainstream phonetic variants. Mean values are presented per word, aggregated across speaker, and across all /oʊ/ tokens (grand mean). These are based on point estimates of non-normalized formant measurements to provide a rough idea of the acoustic profile of each variant condition.

Word	Mainstream Variant (fricated)	Non-mainstream Variant (stopped)
fathered	1.42	4.92
featherweight	1.42	4.83
heathers	1.17	4.58
lathers	1.42	4.50
leathery	1.00	4.50
mothership	1.58	5.00
othered	1.17	5.00
otherness	1.50	4.83
weathered	1.00	4.17
worthiest	1.33	4.58
Grand Mean	1.30	4.69

Table 6.2: /ð/ auditory stimuli phonetic profile. Mean ratings on a 5-point Likert scale based on auditory judgments of how clearly /ð/ was fricated or flapped per word (1=clearly fricated, 5=clearly flapped). Mean values are presented per word, aggregated across speaker, and across all /ð/ tokens (grand mean). Rating values for each individual token are first calculated as the mean of the two coders' ratings.

Variable	Var.	Target Speakers						Filler Speakers					
		Sp1	Sp2	Sp5	Sp8	Sp10	Sp11	Sp3	Sp4	Sp6	Sp7	Sp9	*Total*
/oʊ/-backing	M	10	10	10	10	10	10	0	0	0	0	0	60
	N	10	10	10	10	10	10	0	0	0	0	0	60
/ð/-stopping	M	10	10	10	10	10	10	0	0	0	0	0	60
	N	10	10	10	10	10	10	0	0	0	0	0	60
PIN-PEN	M	10	10	10	9	10	10	0	0	0	0	0	59
	N	10	10	10	9	10	10	0	0	0	0	0	59
FEEL-FILL	M	10	10	10	10	10	10	0	0	0	0	0	60
	N	10	10	10	10	10	10	0	0	0	0	0	60
/u/-fronting	M	10	10	10	10	10	10	0	10	8	0	10	88
	N	10	10	10	10	10	10	0	10	8	0	10	88
/æN/-raising	M	10	10	10	10	10	10	0	10	10	9	0	89
	N	10	10	10	10	10	10	0	10	10	9	0	89
/æ/-backing	M	10	10	10	9	10	10	9	0	0	7	7	82
	N	10	10	10	9	10	10	9	0	0	7	7	82
/t/-deletion	M	10	10	10	10	10	10	10	10	0	9	0	89
	N	10	10	10	10	10	10	10	10	0	9	0	89
	Total	160	160	160	156	160	160	38	60	36	50	34	1174

Table 6.3: Token count for each auditory stimulus condition per speaker.

speaker-variable combination because no tokens of the tenth word were judged as usable.

Five other speakers were designated as filler speakers (listed in Table 6.3). These included speakers who grew up in the US, Canada, or Asia, involving a range of racioethnic impressions including the two speakers identified as most likely Asian and most likely white. Tokens selected for each filler speaker included only certain filler variables (see Table 6.3). In the end, each of the four filler variables were represented by three different filler speakers while each speaker's voice was distributed across two or three variables. The number of tokens included for each speaker varied between seven to ten words per variable. The majority of exclusions were made based on unsuccessful target pronunciations or poor sound quality in parts of the recordings. In total, the full set of auditory stimuli consisted of 1174 unique tokens across all variables, words, speakers, and variants (Table 6.3).

6.3.2 Social-Visual Stimuli

Critical visual stimuli consisted of three sets of three photos that were used to represent the targeted social personae. These personae were labeled as: Ethnic Asian (EA), Mainstream Asian (MA), and Mainstream White (MW). For a detailed description of the process of selecting photos to represent each persona as well as the perceived social characteristics of each, see the norming surveys in Chapter V. The photo groups used in the current experiment are taken from the visual stimuli used in Survey 2 (described in Section 5.5.1.2).[6]

Personae were designed to contrast on two social dimensions, namely race/ ethnicity (Asian vs. White) and cultural alignment (Ethnic vs. Mainstream). The Ethnic Asian persona was conceptualized around the "FOB" character, an Asian American that is more strongly oriented to Asian culture than American culture. This persona may be linked to the use of Asian-associated hair (e.g., sleek and styled), makeup (e.g., straight brows), fashion (e.g., jumper dress), and speech styles (e.g., foreign-accented English). The Mainstream Asian persona was modeled after a "normal" or bicultural Asian American character—one who grew up in the US, has American-style appearance (e.g., casual ponytail, angled brows, tank top) and speech patterns (e.g., American-accented English), and is aligned with both American culture and Asian culture to some extent. Finally, the Mainstream White persona was intended to elicit impressions of a white American who grew up in the US, is closely connected with American culture, and presents themselves with American dress and speech styles.

An additional set of three photos representing a filler persona, labeled Ethnic White (EW), was included to balance perceived race/ethnicity over the course of

[6]Stimuli are available from the author on request. I elected to not publish the stimuli publicly because the photos depict identifiable individuals and this research further involves categorizing them into particular social identities which (a) may or may not match their own identities, and (b) may be considered a stigmatized or negatively perceived identity.

the task as well as to increase perceived social variation within the White photos. In this context, the Ethnic White category represented people who were perceived as racially White but were also perceived as less likely to be from the US compared to the Mainstream White persona. I chose to label the group as "Ethnic" due to conceptual similarity with Ethnic Asian in terms of being perceived as less American. Importantly, however, this filler persona is not meant to be grouped with Ethnic Asian along the dimension of cultural alignment as the interpretations of cultural orientation for each racioethnic group and the degree of difference from their Mainstream persona counterpart are not comparable.

As three photos were included per persona, this resulted in 12 social-visual stimulus items in total. To create the final stimuli, stock photos were cropped to the face and shoulders with the same aspect ratio and uniformly resized. Then, they were normed for various social characteristics using a series of social evaluation surveys that included both open-ended text responses and 7-point semantic differential scale ratings (as reported in Chapter V; see Section 5.4.2 and Section 5.5.1.2). The four persona photo sets were constructed based on these responses and ratings (Table 6.4). Specific photos were selected for inclusion after ensuring that all photos within a persona condition elicited similar perceptual impressions that aligned with the characteristics of that persona as conceptualized above; general alignment with the social characteristics of the voices in the auditory stimuli (e.g., age, ethnicity) were also taken into account. Photos were otherwise ensured to be as well-matched as possible on non-target social characteristics across conditions, meaning that photos associated with particularly extreme or unique social impressions were excluded.

All individuals depicted in the stimulus photos were typically perceived as women in their twenties or thirties. When asked to guess the person's ethnic background, photos in the two Asian personae conditions were very likely to be perceived as ethnically Asian, or specifically East or Southeast Asian, while photos in the two White personae were likely to be perceived as white or of European descent. The key variables used to differentiate the personae conditions were four scales relating to perceived "Americanness". Responses to three of these generally patterned together to represent what I interpret as an "American-raised" dimension (American vs. Not American, Native speaker of English vs. Non-native speaker of English, American-accented vs. Foreign-accented). The remaining scale represented a culturally "American-aligned" dimension (Aligned with American culture vs. Aligned with ethnic culture). As seen in Table 6.4, photos categorized as Ethnic Asian were generally rated as low on both "American-raised" and "American-aligned" dimensions while Mainstream White was rated as very high

Scale	**EA**	**MA**	**MW**	**EW**
American	-0.83 (0.13)	0.41 (0.08)	0.95 (0.26)	0.04 (0.18)
Native speaker of English	-1.02 (0.20)	0.38 (0.12)	0.97 (0.25)	0.32 (0.21)
American-accented	-1.11 (0.18)	0.32 (0.13)	0.87 (0.25)	-0.02 (0.12)
Aligned with American culture	-0.93 (0.14)	-0.13 (0.19)	0.96 (0.18)	0.16 (0.11)
Enthusiastic	0.16 (0.14)	0.21 (0.45)	0.32 (0.19)	0.21 (0.05)
Confident	0.04 (0.37)	0.35 (0.49)	0.36 (0.25)	0.34 (0.10)
Friendly	0.43 (0.05)	0.43 (0.10)	0.33 (0.09)	0.33 (0.11)
Likeable	0.44 (0.21)	0.33 (0.22)	0.28 (0.17)	0.22 (0.14)
Attractive	0.24 (0.30)	0.03 (0.08)	0.41 (0.27)	0.17 (0.09)
Intelligent	0.38 (0.20)	0.40 (0.29)	0.07 (0.31)	0.12 (0.10)
Feminine	0.36 (0.36)	0.01 (0.41)	0.75 (0.29)	0.42 (0.07)
Clear speaker	-0.13 (0.27)	0.41 (0.12)	0.55 (0.20)	0.27 (0.11)
Cool	-0.08 (0.13)	-0.08 (0.58)	-0.11 (0.12)	-0.22 (0.05)
Casual	-0.53 (0.50)	-0.20 (0.93)	-0.61 (0.71)	-0.30 (0.44)
Nerdy	-0.34 (0.21)	-0.43 (0.56)	-1.29 (0.29)	-0.89 (0.05)
Slow speaker	-0.61 (0.10)	-0.90 (0.44)	-0.94 (0.26)	-0.86 (0.16)

Table 6.4: Perceived social characteristics of photos representing each persona condition (EA = Ethnic Asian, MA = Mainstream Asian, MW = Mainstream White, EW = Ethnic White), based on social evaluation judgments. Values represent standardized (z-scored) mean ratings (SD) on sixteen 7-point semantic differential scales, average across all responses for the three photos per persona condition (M = 0.03; Min = -1.60; Max = 1.26). Scale labels represent the interpretation of the the maximum positive value.

on both. Photos grouped under Mainstream Asian, on the other hand, were rated as moderately high on "American-raised" while being comparatively lower on "American-aligned". I link this to the bicultural aspect of the Mainstream Asian persona. Finally, the filler photos selected for Ethnic White were rated as varying in the moderate range for all four scales. Overall, these ratings show that social impressions of the visual stimuli pattern in expected ways given the targeted persona constructs.

6.3.3 Sociolinguistic Questionnaire

A modified version of the ideology, social network, and personal demographics survey sections from Chapter IV was used as a post-task questionnaire. Detailed open-ended response questions about Asian American-related ideologies were omitted in favor of more questions that specifically quantified degree of awareness for the target personae and variables of interest (see Appendix D for relevant sections of the survey). The survey was presented to participants in five sections, ordered as follows: Social Ideology, Linguistic Ideology, Social Exposure, Linguistic Exposure, and Personal Demographics.

The first section about social ideology asked what participants thought about "certain cultural ideas relating to Asian Americans". This began with general questions about who they would typically describe as Asian American as well as whether and to what degree they personally identified as Asian American. Then, they were asked a set of nine rating questions about three different Asian American social types; these structure of these questions are described in more detail below. The social types were introduced as "three potential types of Asian Americans that have been described by other Asian Americans". Social type labels and descriptions were constructed based on responses to the ideology survey in Chapter IV and presented in this order: "Bicultural" Asian American (e.g., Normal Asian), "Culturally Asian" Asian American (e.g., FOB Asian), and "Culturally American" Asian American (e.g., Whitewashed Asian).

The second section about linguistic ideologies was structured similarly but asked what participants thought about "how Asian Americans speak." This began with questions asking whether respondents believed there were differences between how some Asian Americans and other American speak, how they would describe an Asian American accent, and whether and how Asian American accented speech differs from Asian foreign accented speech. Next, they answered a similar set of nine questions adapted to target experiences with ideologies about an Asian American accent. This was followed by questions focused on their ideologies about the critical linguistic variables of /oʊ/-backing and /ð/-stopping as relating to Asian American speech. These were introduced as "two potential aspects of an Asian American accent that have been described by other Asian Americans". Again, labels and descriptions were constructed based on responses to the ideology survey in Chapter IV, and /oʊ/-backing was presented before /ð/-stopping. First, respondents were asked if they thought that example words with the target sound would be pronounced differently by Asian Americans compared to White Americans. If they answered "yes", they were asked to explain the pronunciation differences if they could. Then, they were provided with a specific description of the targeted pronunciation variants and and asked to answer the same nine questions. /ð/-stopping was described as pronouncing "the 'th' sound with a 'd' sound" while /oʊ/-backing was described as pronouncing "the 'o' sound as more 'drawn out', 'longer', 'elongated', or 'emphasized'" and also potentially "sounding 'rounder' and 'lower in tone'".

The series of nine questions presented with each targeted ideology was designed to probe degree of conscious awareness, perceived accuracy, associated valence, and personal alignment of to these sociolinguistic ideologies (see Table 6.5

for the questions assessing /oʊ/-backing and /ð/-stopping; exact question wording for the other measured ideologies can be found in Appendix D). These were formatted as 7-point Likert scales. In particular, the first five questions were designed to measure the extent of subjective frequency and familiarity that respondents had with each concept; this aimed to provide an estimate for the construct of awareness, conceptualized as conscious access to certain ideas or beliefs (for additional details of these measures and their interpretations, see Section 6.3.7 on data analysis).

The next two sections focused on sociolinguistic exposure, asking questions about participants' "social and linguistic interactions" as well as their "language experiences". The former section included questions about the percentage of respondents' social network (e.g., family, friends, colleagues) that represented different racial/ethnic backgrounds as well as the percentage of peers that represented Asians of different backgrounds. They were also asked how much time they spent with each of these groups. The latter section about language asked for estimates like the percentage of English they heard spoken by people of different backgrounds. Finally, the last section included questions on their "demographic background and identity". To assess ethnic background, respondents were asked to list up to five identity labels they would use to describe their ethnicity and rate how much they identified with each label on a 7-point scale. They were also asked about their immigrant generation, along with basic information such as age, gender, and locations lived.

6.3.4 Task Design

The experimental task was a two-alternative forced choice (2AFC) speaker identification task using a mouse tracking paradigm (e.g., Spivey et al., 2005; Blazej & Cohen-Goldberg, 2015; Lin et al., 2022). On each trial, listeners would

Construct	Scale	Question Text
Awareness	Frequency	How often does the idea that some Asian Americans pronounce [”o” with more “emphasis” / ”th” more like ”d”] come up in your daily experiences? (e.g., hearing about it in conversations, reading about it in online media, noticing it in your social interactions, thinking about it)
	Familiarity	How familiar is the idea that some Asian Americans pronounce [”o” with more “emphasis” / ”th” more like ”d”] to you?
	FamiliarityTime	When did you become familiar with the idea that some Asian Americans pronounce [”o” with more “emphasis” / ”th” more like ”d”]?
	PersonalExamples	Think about the people that you hear speaking in your personal life (e.g., social networks, local community). How often do you feel you notice examples of Asian Americans pronouncing [”o” with more “emphasis” / ”th” more like ”d”]?
	MediaExamples	Think about the people that you hear speaking in media (e.g., YouTube, Instagram, TV shows, movies). How often do you feel you notice examples of Asian Americans pronouncing [”o” with more “emphasis” / ”th” more like ”d”]?
Accuracy	Accuracy	How realistic, accurate, or true do you think the idea that some Asian Americans pronounce [”o” with more “emphasis” / ”th” more like ”d”] is?
Valence	Valence	Do you consider the idea that some Asian Americans pronounce [”o” with more “emphasis” / ”th” more like ”d”] to be positive, negative, or neutral?
Alignment	OtherAscription	How often have other people told you that you pronounce [”o” with more “emphasis” / ”th” more like ”d”]?
	SelfAscription	To what extent do you feel like you might pronounce [”o” with more “emphasis” / ”th” more like ”d”]?

Table 6.5: Survey questions assessing experience with ideologies about /oʊ/-backing and /ð/-stopping as sociophonetic variables associated with Asian American speech. This includes nine individual 7-point Likert scale ratings questions targeting four constructs: AWARENESS, ACCURACY, VALENCE, and ALIGNMENT.

hear a word with either a Mainstream or Non-mainstream pronunciation. Then, they would choose who the speaker was from two of the personae options (Mainstream Asian, Ethnic Asian, Mainstream White, Ethnic White) represented by a pair of photos. This experiment was designed with mouse tracking primarily to make it comparable in design to a planned follow-up experiment assessing decision making during real-time word processing. While a secondary goal was to explore whether mouse tracking measures could be used to assess decision making during a speaker identification task where there is no correct answer, this method is not central to answering the current research questions. Thus, I do not discuss mouse tracking for this experiment except to describe the task procedure.

In detail, the task involved a within-subjects design that crossed two phonetic variant conditions, represented by the audio stimuli, and four persona pairing conditions, represented by the visual stimuli. These conditions were presented for all eight linguistic variable conditions: two critical variables (/oʊ/-backing, /ð/-stopping), two comparison variables (PIN-PEN merger, FEEL-FILL merger) and four filler variables (/u/-fronting, /æ/-backing, /æN/-raising, /t/-deletion). As such, each trial represented a combination of a linguistic variable, phonetic variant, and persona pairing, resulting in 64 unique conditions (8 variables $\times$ 2 variants $\times$ 4 pairings). This design enables assessment of how a particular variant may be socially interpreted across multiple persona pairing contexts per individual.

Condition trials that included the four critical and comparison variables were presented three times each. The specific auditory stimulus that participants heard per trial were randomly selected from the relevant pool of recordings. This pool would include all words produced by the six target speakers that contained a particular variable and was pronounced with a particular variant. In these trials, the critical persona pairings were the three pair combinations of EA, MA, and MW (i.e., EA-MA, MA-MW, EA-MW). A fourth condition of EW-MW contained no

Asian photo option, serving as a control. The specific visual stimulus (i.e., pair of images representing the two personae) that participants saw were randomly selected from all combinations of the three photos representing each persona placed on either side of the screen.

Filler condition trials included the four filler variables and were presented twice, once in a target voice and once in a filler voice. Only two repetitions were included to minimize the length of the experiment. In order to present all personae relatively evenly across the experiment, filler trials presented EW instead of MW for pairings comparing Asian and White personae. Again, specific audio and visual stimuli were randomly selected from the relevant pool of recordings and photo combinations per trial.

In total, the experiment included 160 trials of which 24 were /oʊ/-backing and 24 were /ð/-stopping trials (3 trials per phonetic variant-persona pairing condition). Target variables (n = 96) were encountered slightly more often than filler variables (n = 64). The majority of trials involved auditory stimuli produced by the target voices (n = 128), interspersed with trials including filler voices (n = 32). Certain personae and personae combinations appeared more than others: EA-MA and EW-MW were the most common due to appearing with both target and filler variables (n = 40), followed by the pairings unique to target variables (EA-MW and MA-MW; n = 24) and those unique to filler variables (EA-EW and MA-EW; n = 16). However, the result was a balanced presentation of Asian- and White-perceived faces where half the trials paired photos of the same race/ethnicity and half paired photos of contrasting race/ethnicity.

For all participants, the first five trials were randomly selected from filler voice conditions, serving as practice trials. The remaining trials were randomly ordered. Although the number of each trial condition was fixed, the number of specific speakers, words, photos, and photo locations were variable across participants.

Trial Type	Variable	Variant	EA-MW	MA-MW	EA-MA	EW-MW	EA-EW	MA-EW	*Total*
Target Speaker + Target Variable	/oʊ/-backing	M	3	3	3	3	0	0	12
		N	3	3	3	3	0	0	12
	/ð/-stopping	M	3	3	3	3	0	0	12
		N	3	3	3	3	0	0	12
	FEEL-FILL	M	3	3	3	3	0	0	12
		N	3	3	3	3	0	0	12
	PIN-PEN	M	3	3	3	3	0	0	12
		N	3	3	3	3	0	0	12
Target Speaker + Filler Variable	/æ/-backing	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
	/æN/-raising	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
	/t/-deletion	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
	/u/-fronting	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
Filler Speaker + Filler Variable	/æ/-backing	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
	æN/-raising	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
	/t/-deletion	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
	/u/-fronting	M	0	0	1	1	1	1	4
		N	0	0	1	1	1	1	4
		Total	24	24	40	40	16	16	160

Table 6.6: Trial count for each auditory and visual stimulus condition.

Two restrictions were placed on the selection of specific stimuli used in a particular trial: First, no auditory stimulus (i.e., speaker-word combination) could be presented more than once per participant. Second, consecutive trials could not contain the same speaker, word, or images (on either side). After these restrictions were applied, a stimulus combination would be randomly selected from the remaining options. In this way, participants would be presented with 160 unique recordings involving a mix of voices, words, and photo options (Table 6.6).

6.3.5 Procedure

Interested individuals first completed a screening survey that assessed the inclusion criteria of whether they (a) considered themselves ethnically Southeast or East Asian; (b) grew up or lived for a significant amount of time in California, Oregon, or Washington state;[7] (c) were between the ages of 18-45; and (d) currently lived in the US. Word-of-mouth advertising (i.e., emails, flyers) explicitly asked individuals of this demographic to volunteer. Participants recruited on Prolific (www.prolific.com), a participant recruitment website, were internally filtered using built-in screeners for age, ethnicity, and the US state they resided in.

The inclusion criteria were designed to narrow in on a specific population of Asian Americans that met the needs of this study. First, due to the larger concentration of Asian Americans in California and the West Coast more generally, these participants were overall expected to be most familiar with the targeted social and linguistic ideologies under study, which would allow for a range of awareness levels. In addition, due to having relatively similar social identities to those represented by the voices and photos in the experimental stimuli (e.g., ethnicity, age, residential history), these participants were expected to interpret the people they saw and heard in the experiment in a similar manner (e.g., as peers).

Individuals who passed the screening survey were invited to participate with instructions and a URL to access the study in a web browser (either via email or Prolific). They were told that the study was about racial/ethnic stereotypes and social speech perception, but not that it would focus on AACs. They were also informed that the task would require headphones and that they should complete the task in a quiet location free of distractions.

Participants completed the study remotely through a web browser on a desktop

[7]Potential participants from Oregon and Washington states were included in the recruitment process in case sufficient numbers of Californian participants were not attained. In practice, this was not required and only participants from California were invited to participate.

or laptop computer (i.e., no participants used mobile devices like smartphones or tablets). Upon accessing the study URL, participants were directed to a Qualtrics form where they reported the equipment they would use, namely the type of mouse (e.g., wired, wireless, or trackpad) and headphones (wired or wireless, in-ear or over-ear). Then, they read the informed consent form and provided their consent before continuing. After being redirected to the experiment, they completed a binaural tone test as a headphone check (Woods et al., 2017), for which they had to pass at least 5 out of 6 trials to proceed. Those who failed the test three times were prevented from continuing the study.

The main speaker identification task was presented next. Participants were told that they would hear words read by multiple different people and their goal was to select the person they think is speaking from two photos as quickly as possible. They were informed that each person was recorded reading the same set of words and asked to make choices based on how they spoke instead of what they were saying. Finally, they were told that to facilitate a quick response, they should start moving their mouse immediately after clicking the button to play the recording, even before coming to any decisions.

At the beginning of each trial, participants were presented with a word in the middle center of the screen and two photos on the top left and right of the screen. After 500 ms, a "START" button would appear at the bottom center of the screen. When participants clicked on 'START', the button would disappear while the word and photos remained in place. A recording of the word would play after a 500 ms delay.[8] The trial ended once the participant clicked on either of the two photo response options. If no valid mouse click was detected within five seconds after the start of the auditory stimulus, the trial would end automatically. Regardless,

[8] The delay is meant to given participants a chance to begin moving their mouse earlier in order to better detect mouse movements during decision making, following procedures in Spivey et al. (2005).

each trial was separated by a 500 ms long blank screen. In total, there were 160 trials presented in ten blocks of 16.

After completing the main task, participants were redirected back to Qualtrics to answer a few questions reflecting on the task followed by the sociolinguistic questionnaire. Finally, they were debriefed about the purpose of the experiment and provided an opportunity to withdraw their consent in light of this information. In the end, no participants elected to withdraw consent. The full study generally took under one hour to complete. All who completed the study received $12 in compensation.

6.3.6 Participants

In total, 63 Asian Americans completed the study, recruited via Prolific (n = 45) and word-of-mouth (n = 18). After excluding 3 Prolific participants (2 for not following instructions; 1 for failing to meet recruitment criteria), 60 were retained for analysis. Ages in this sample span the full inclusion range of 18-45 (M = 28.7, SD = 6.6). Gender representation was relatively balanced, including 33 women (55%), 26 men (43%), and 1 non-binary individual (2%). A majority of participants reported at least some college education ("some college" = 8%, "4-year degree" = 65% , and "doctorate" = 5%). In terms of residential history, most (83.3%) reported that they had grown up in California, lived there for the majority of their life, and still lived there. Regardless, all had lived in California for at least 9 years total.[9]

In terms of ethnicity, all participants self-identified as East or Southeast Asian. Degree of alignment with "Asian American" as an identity label was very high overall (based on a 7-point Likert scale from 1 = *not at all* to 7 = *completely*; *Median* = 7, M = 6.23, SD = 1.16). Table 6.7 reports the specific ethnicity breakdown,

[9]This arbitrary threshold was based on the minimum number of years one would have likely lived in a location in order to report spending the majority of their childhood years there (i.e., before age 18).

Ethnicity (coded)	N	Ethnicity (coded)	N
Chinese	16 (26)	Chinese-Filipino	1 (1)
Vietnamese	13 (11)	Laotian	1 (1)
Filipino	11 (11)	Multiracial	1 (1)
Taiwanese	8 (4)	Hong Konger	1 (0)
Korean	3 (3)	Singaporean	1 (0)
Japanese	2 (1)	Taishanese	1 (0)
Cambodian	1 (1)		

Table 6.7: Participant Ethnicity. Counts for specific ethnic identities coded from self-reported labels that participants rated as the one(s) they most strongly identified with. To provide a summary of the diversity in the sample, the non-Chinese label was used when Chinese and another ethnic identity were equally rated. For context, counts using the Chinese label instead are reported in parentheses.

coded based on self-reported labels that participants most strongly identified with. The most commonly represented ethnic groups in this sample include Chinese, Vietnamese, Filipino, and Taiwanese. The majority (75%) identified as second-generation (i.e., born and raised in North America with a first-generation immigrant parent). The sample also included representation from Asian Americans who identified as first generation (8.3%), 1.5 generation (8.3%), and third generation[10] and above (8.3%).

Based on social network questions focused on race/ethnicity, this sample included individuals with a wide range of exposure to racial diversity and co-ethnics. One estimate of racial diversity is the Diversity Index (DI), which represents the probability that two people chosen at random will be from different racioethnic groups (E. Jensen et al., 2021).[11] Based on the percentage of time participants reported typically spending with ten different racioethnic groups, the average DI of the sample is moderate but the distribution shows that there was a wide range of experiences from interaction with only one racioethnic group to regular interaction with people of various backgrounds (M = 0.50, SD = 0.25,

[10] This includes one individual who identified as both second and third generation.

[11] Zero indicates that everyone in the population has the same racioethnic characteristics, while a value close to 1 indicates that everyone in the population has different characteristics.

range: 0–0.82). Exposure to East or Southeast Asians was relatively high, but again showed large variation within the sample based on the number of different people they closely interact with (*M* = 0.60, *SD* = 0.25, range: 0.1–1) and consistently interact with (*M* = 0.49, *SD* = 0.28, range: 0–1), as well as the total time spent with East or Southeast Asians (*M* = 0.53, *SD* = 0.27, range: 0–1).

Considering exposure to peers of different backgrounds, the majority of participants spent the most time interacting with Asians who grew up in North America (*M* = 0.49, *SD* = 0.25) and non-Asians (*M* = 0.36, *SD* = 0.24), rather than Asians who grew up in Asia (*M* = 0.13, *SD* = 0.17). Accordingly, of their time being exposed to English, they heard English the most from Asian Americans who grew up in North America (*M* = 0.33, *SD* = 0.20) and White Americans (*M* = 0.30, *SD* = 0.19), followed by Latin/Hispanic Americans (*M* = 0.17, *SD* = 0.17), Asian Americans who grew up in Asia (*M* = 0.11, *SD* = 0.11) and Black Americans (*M* = 0.09, *SD* = 0.09). Overall, these metrics indicate that the present sample of Asian Americans from California represented people from a variety of sociolinguistic contexts.

6.3.7 Data Analysis

6.3.7.1 Ideological Self-Report

The data from the sociolinguistic questionnaire involved all 60 participants' responses. Descriptive statistics are provided for three sets of nine 7-point Likert scale ratings asking about Asian American speech, /oʊ/-backing, and /ð/-stopping (see full questions in Appendix D). The nine rating scales are discussed both independently and in terms of the constructs they were intended to measure (see list of scales in Table 6.5). The four constructs include composite values for AWARENESS and ALIGNMENT along with single-scale values for ACCURACY and VALENCE; these are described in more detail below.

Five scales (FREQUENCY, FAMILIARITY, FAMILIARITYTIME, PERSONALEXAM-

PLES, MEDIAEXAMPLES) were designed to assess awareness via gradient measures of subjective frequency and familiarity as associated with each concept. To obtain a single numerical value representing an individual's degree of awareness, a composite AWARENESS score was calculated as the average rating of those five scales for each ideology. A score of one is taken to represent "not at all aware" while a score of seven would represent "extremely aware". Separately, ratings for ACCURACY represented the extent to which respondents believed the construct to be true (1 = *not at all true* to 7 = *extremely true*). The VALENCE scale was used to estimate attitudes in terms of how positively or negatively each concept was viewed (1 = *extremely negative* to 7 = *extremely positive*). Finally, the last two scales (OTHERASCRIPTION, SELFASCRIPTION) were averaged to output an ALIGNMENT score, interpreted as the extent to which respondents were believed to produce an "Asian American accent", /oʊ/-backing, or /ð/-stopping. A score of one is taken to represent "not at all aligned" while a score of seven would represent "completely aligned".

6.3.7.2 Photo Selections

To probe whether the presence of /ð/-stopping or /oʊ/-backing can influence AAC speaker identification behavior, I assess how likely listeners were to select the Ethnic (E), Asian (A), or Ethnic Asian (EA) photo when they heard Nonmainstream (N) variants compared to Mainstream (M) variants. The proportion of these target selections is referred to as PropEA, and the difference in PropEA between variant conditions (PropEA$_{N-M}$) is examined within each persona pairing condition. The target pairings contain at least one Asian persona option (i.e., MA-MW, EA-MA, EA-MW) while the white-only condition (EW-MW) serves as a control. Target pairings are evaluated separately because each has a different interpretation: MA-MW contrasts race/ethnicity within culture (Asian vs. White), EA-MA contrasts culture within race/ethnicity (Ethnic vs. Mainstream), and EA-MW contrasts both race/ethnicity and culture. Regardless, a positive value of PropEA_{N-M} would be interpreted as a greater likelihood of selecting the Ethnic and/or Asian photo when hearing the target Non-mainstream variant (/ð/-stopping or /oʊ/-backing), which is the predicted effect if the target phonetic features are linked to AACs generally or tied to specific AAC personae. In total, 1432 trials were analyzed for the ð/-stopping variable condition (excluding 8 null trials where no response was provided) and 1434 trials were analyzed for the/oʊ/-backing variable condition (excluding 6 null trials where no response was provided).

Analyses were conducted via Bayesian hierarchical mixed-effects logistic regression using *brms* (Bürkner, 2017) in R (R Core Team, 2020). Separate models were fit for the two target variables, /ð/-stopping and /oʊ/-backing.[12] The dependent variable was the binary-coded mouse click response for the more Ethnic and/or Asian photo option (EA option = 1, non-EA option = 0). Fixed effects included VARIANT (2 levels: N, M), PERSONAPAIRING (4 levels: EW-MW, MA-MW, EA-MA, EA-MW), and the VARIANT×PERSONAPAIRING interaction. All input variables were sum coded. The model also included by-participant, by-speaker, and by-word random intercepts and slopes for VARIANT× PERSONAPAIRING.[13]

The models used weakly informative priors for all population-level effects, defined as normal distributions centered on 0 with a standard deviation of 5 (in log-odds space), or Normal(μ=0, σ=5). Priors for population-level standard deviations were Student-t(μ=0, σ=2.5, ν=0) distributions while priors for random effect correlations were LKJ distributions where η was 1. All models were fit with 4 chains and

[12]Model syntax: $brm(mouseClick.EA \sim variant * personaPairing + (1 + variant * personaPairing|participantID) + (1 + variant * personaPairing|speaker) + (1 + variant * personaPairing|word), family = \text{'bernoulli'})$

[13]A more complex version of this model was also fit. This model included an additional random intercept and random slopes for the photo pair that participants saw per trial. However, it resulted in model convergence difficulties. Because parameter estimates from the more complex model were comparable to the simpler model, I chose to report the estimates using the more stable model.

2,000 iterations (1,000 warm-up). Estimated marginal means were extracted using *emmeans* (Lenth, 2023). As the key contrast of interest was the difference between VARIANT within each PERSONAPAIRING, only pairwise comparisons of VARIANT (N-M) nested within PERSONAPAIRING are presented.

To summarize the model output, I report median posterior point estimates for the key N-M VARIANT difference effect per PERSONAPAIRING, along with the corresponding 95% percentile credible intervals, and the posterior probability that the estimate is greater than zero. The last value represents the probability that the key effect is in the positive direction (see Nicenboim & Vasishth, 2016). To interpret the results, I qualitatively assess the strength of evidence. I describe evidence for an effect as strong if the 95% credible interval excludes zero while also considering moderate-to-weak evidence for effects based on the degree of certainty indicated by the credible intervals and the probability of a positive effect.

6.3.7.3 Awareness

To assess the role of awareness in mediating speaker identification behavior, analyses were again conducted using hierarchical mixed-effects logistic regression using *brms* (Bürkner, 2017) in R (R Core Team, 2020). Two models were fit, one each for /ð/-stopping and /oʊ/-backing; these were constructed based on the previous analysis with the additional inclusion of VARIANTAWARENESS as a covariate.[14] VARIANTAWARENESS was the standardized (z-scored) /oʊ/-backing and /ð/-stopping composite AWARENESS scores, calculated as the unweighted average of the five awareness-related questions in the self-report data. Models included the same weakly informative priors, including a Normal distribution centered on zero with a standard deviation of 5 for population-level effects. They were fit with

[14]Model syntax: $brm(mouseClick.EA \sim variant * personaPairing * variantAwareness + (1 + variant * personaPairing|participantID) + (1 + variant * personaPairing * variantAwareness|speaker) + (1 + variant * personaPairing * variantAwareness|word), family = \text{'}bernoulli\text{'})$

four chains and 2,000 iterations (1,000 warm-up). Estimated marginal means of linear trends were extracted using the *emtrends* command from the *emmeans* package (Lenth, 2023). The same method of summarizing and interpreting results are applied for this analysis.

6.3.8 Predictions

One aim of this study was to assess the extent of awareness for /oʊ/-backing and /ð/-stopping as ideological features of AAC speech. Based on self-report measures, I predict that both /oʊ/-backing and /ð/-stopping will be ideologically associated with Asian (American) identity at the community level, such that at least a portion of respondents will report some degree of awareness of these features in Asian American speech. Second, based on the results of Chapter IV, I predict that /ð/-stopping will be linked to higher awareness scores compared to /oʊ/-backing. Regardless, I expect high individual variation in awareness scores reported across respondents.

To address whether the presence of /oʊ/-backing and /ð/-stopping influences perceptual behavior in the speaker identification task, the main outcome of interest was photo response selections. If listeners associate /oʊ/-backing and /ð/-stopping variants more with Asian (American) identity compared to a White (American) identity—and if they are perceptually sensitive to these cues—then they should be more likely to select an Ethnic Asian or Mainstream Asian photo that is presented opposite to a Mainstream White photo when they hear /oʊ/ and /ð/ words pronounced with /oʊ/-backing and /ð/-stopping. In other words, I predict that in persona pairing conditions contrasting Asian and White race/ethnicity (EA-MW and MA-MW), the proportion of Asian photos selected as the speaker in the Non-Mainstream variant condition will be greater than the proportion selected in the Mainstream variant condition.

Similarly, if listeners more strongly associate these critical variables with an Ethnic (Asian) cultural alignment compared to a Mainstream (American) cultural alignment, they should be more likely to select an Ethnic Asian photo as the speaker, compared to Mainstream Asian or Mainstream White, when they hear /oʊ/-backing and /ð/-stopping. That is, I predict that in critical persona pairing conditions contrasting Ethnic and Mainstream cultural alignment (EA-MA and EA-MW), the proportion of Ethnic Asian photos selected in the Non-Mainstream variant condition will be greater than the proportion selected in the Mainstream variant condition.

In the case that /oʊ/-backing and /ð/-stopping have both race/ethnicity and cultural alignment social associations, different patterns could emerge based on the strength of associations with each dimension (or persona) and how effects are mediated by multiple social associations. For example, I would expect in this scenario that the persona pairing condition contrasting both race/ethnicity and cultural alignment, namely EA-MW, will result in the largest predicted effect. In general, if different behavioral patterns are found in response to Mainstream Asian versus Ethnic Asian personae, indicating that cultural alignment within the macrosocial category of "Asian" is a relevant microsocial dimension for these variables, this provides support for persona-based sociophonetic representations and processing.

In terms of whether awareness modulates strength of sociophonetic associations and therefore social activation during speech perception, higher awareness for /oʊ/-backing and /ð/-stopping is generally predicted to increase the behavioral difference effect between Non-mainstream and Mainstream variant conditions. Examined at the level of community awareness, if I find that one variant is self-reported with a higher average degree of awareness (e.g., /ð/-stopping), that feature should be associated with greater (Ethnic) Asian selection bias in the

Non-mainstream variant condition compared to the other feature. Examined at the level of individual awareness, I similarly predict that individuals who self-report a higher degree of awareness will show greater (Ethnic) Asian selection bias in the Non-mainstream variant condition compared to individuals who self-report lower degrees of awareness.

6.4 Results: Ideological Self-Report

In this section, I describe the awareness of the relevant AAC linguistic ideologies in this sample of Asian Americans from California, based on the results of the sociolinguistic questionnaire.

6.4.1 Asian American Accent

In order to contextualize self-reported ideology and awareness results for /ð/-stopping and /oʊ/-backing, I first report on what this sample of Californian Asian Americans think about the general idea of an Asian American accent. Mean values for the four construct scores are depicted in Figure 6.1. To additionally visualize individual variation and response patterns, data distributions for each of the nine scales are presented in Figure 6.2.

Out of the 60 total respondents, 55 reported believing that *some* Asian Americans speak differently from other Americans (“yes” = 91.7%; “no” = 6.7%; “other” = 1.7%). The same proportion (91.6%) also believed that “Asian American-accented speech” was different from “Asian foreign-accented speech”, making clear that participants were discussing a speech style specific to local Asian Americans.

To flesh out respondents’ beliefs about this local Asian American accent, I first turn to ratings of perceived ACCURACY and VALENCE. For ACCURACY, the sample generally considered this concept to be relatively accurate (*Median* = 5, *M* = 4.83, *SD* = 1.43). In fact, all respondents reported believing that the accent was true to some extent, though there was variation in the level of endorsement. While the most common response was 5 (*quite true*; 38.3%), a sizable number selected 7 (*extremely true*; 18.3%). In terms of VALENCE, the Asian American accent was neither considered positive nor negative in general (*Median* = 4, *M* = 4.12, *SD* = 0.98), with 4 (*neutral*; 71.7%) being the most popular response by far. This suggests that the idea of an Asian American accent is widespread and seen as a neutral fact of Asian American speech.

Next, I consider whether respondents may consider themselves to speak with an Asian American accent. Their ALIGNMENT to an Asian American accent appeared to be relatively low on average (*Median* = 2, *M* = 2.60, *SD* = 1.45). When asked how often they are told they "sound Asian" (OTHERASCRIPTION), most respondents selected 1 (*never*, 45.0%) or 2 (*rarely*, 21.7%). In terms of their own judg-

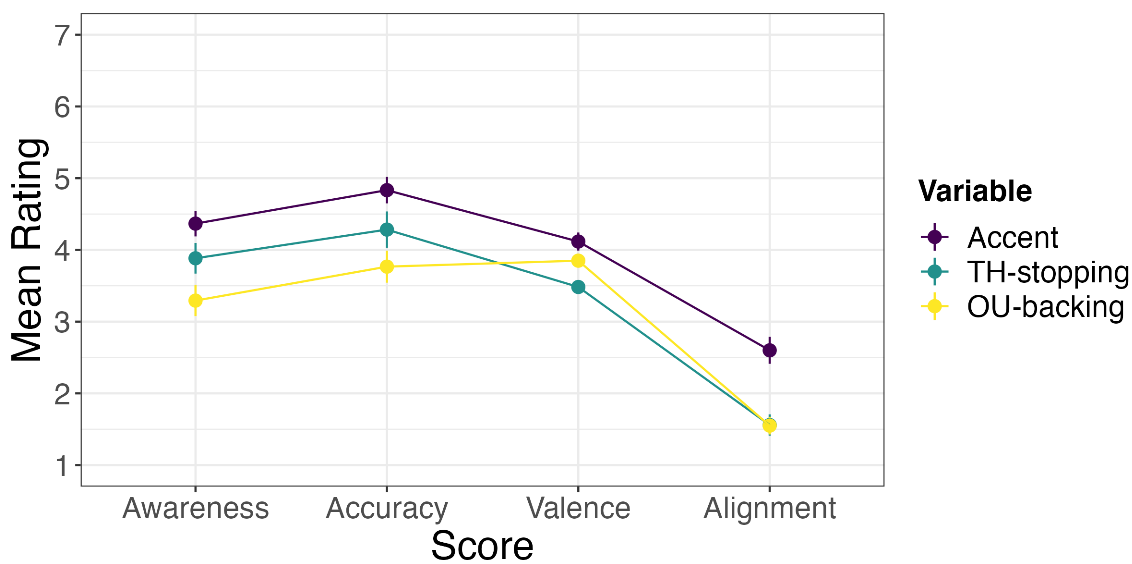

Figure 6.1: Parallel coordinates plot representing the mean value (error bars indicate standard error) per construct and variable being rated. Construct summary scores include multi-scale composite values for AWARENESS and ALIGNMENT along with single-scale values for ACCURACY and VALENCE.

ments on the extent to which they "sound Asian" (SELFASCRIPTION), the most common answer was also 1 (*not at all*; 25.0%). However, the distribution was much more diffuse across the scale, including many selecting 2 (*slightly*; 20.0%) and 3 (*somewhat*; 21.7%). In total, three-quarters of the respondents self-identified as speaking with an Asian American accent to some extent.

While the concept of an Asian American accent was evidently accessible in consciousness, individuals varied as to the degree of awareness they reported. The composite AWARENESS scores indicated that there was moderate awareness on average (*Median* = 4.4, *M* = 4.37, *SD* = 1.38). However, scores were relatively evenly distributed across the full possible range, meaning that individuals varied greatly overall. For example, one participant scored 1, indicting no awareness of the concept of an Asian American accent at all, while four scored 7, indicating they they were extremely aware. Considering responses for each AWARENESS-related scale individually, the observation of overall variability holds: All five questions elicited at least one response for each of the scale steps.

In general, we see three patterns across the five scales (see Figure 6.2). First, regarding perceived FREQUENCY, the idea of an Asian American accent tended to come up in daily experiences relatively infrequently: The top response was 2 (*rarely*; 28.3%). However, respondents subjectively felt that the idea was very familiar. When asked *how* familiar the Asian American accent idea was (FAMLIARITY), the most common response was 7 (*extremely familiar*; 28.3%), followed by 4 (*moderately familiar*; 20.0%) and 5 (*quite familiar*; 18.3%). The overall response pattern was similar when asked *how long* they had been familiar with this concept (FAMILIARITYTIME) with the majority again responding 7 (*as long as I can remember*; 30%), followed by 4 (*some time ago*; 25.0%) and 5 (*somewhat early on*; 21.7%).

In addition, a large portion of the sample recalls encountering the accent in

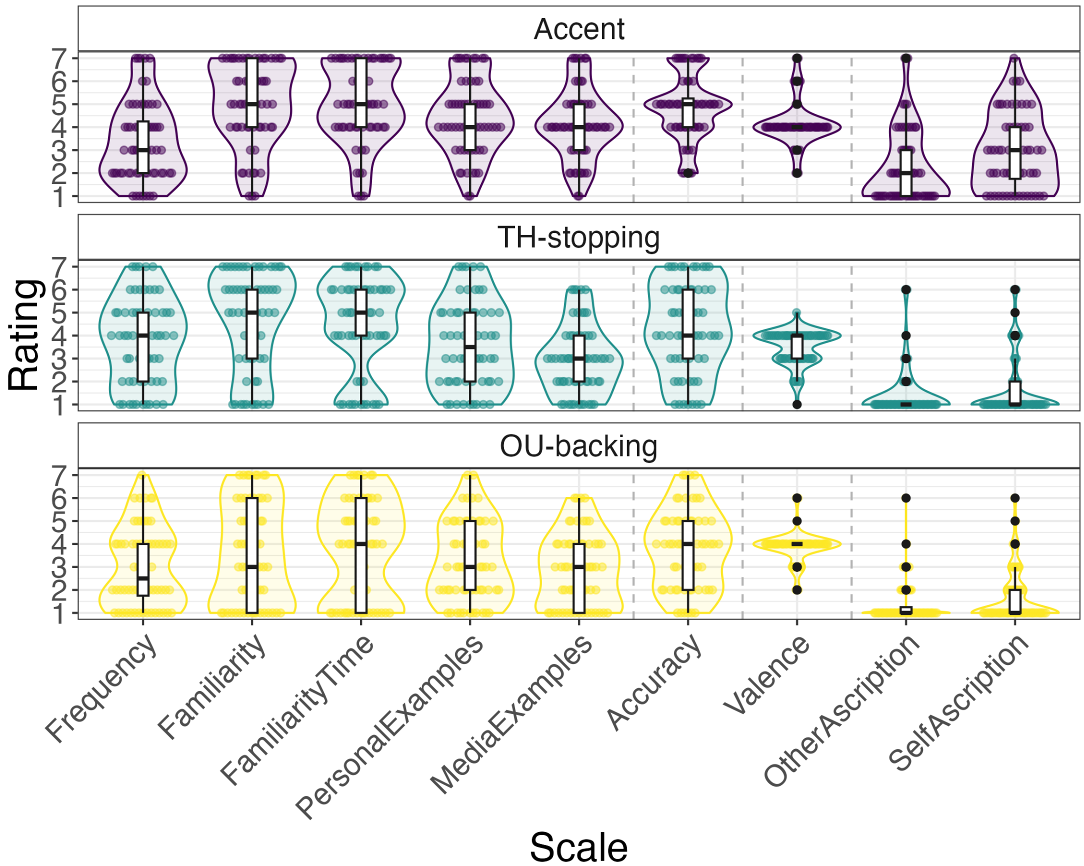

Figure 6.2: Violin and box plots representing the distribution of responses per question and variable being rated. The five scales on the left targeted the construct of AWARENESS while the two scales on the right targeted ALIGNMENT towards the variable in question.

their daily life. Respondents reported noticing examples of an Asian American accent both in their personal life (PERSONALEXAMPLES), where the most common responses were 4 (*sometimes*; 28.3%) and 5 (*commonly*; 21.7%), as well as in media (MEDIAEXAMPLES), where the most common response was 4 (*sometimes*; 40%). These results indicate that while the concept of an Asian American accent is not explicitly discussed very often, many respondents do notice examples of the accent in both interactions and media and it is generally an idea that has felt familiar to many respondents for quite some time.

Taken together, this sample of Asian Americans overwhelmingly had access to the idea of a neutral Asian American accent specific to local Asian Americans who grew up in North America—not just one that is conflated with a foreign accent—and many believed that they used it themselves. On the other hand, the strength of this idea in their consciousness was highly variable across individual experiences, averaging out to moderate awareness. Still, more often than not, an "Asian American accent" was considered to be a familiar concept.

6.4.2 Phonetic Variables

Next, I move on to the results of the targeted AAC variables, /ð/-stopping and /oʊ/-backing, in terms of associated ideologies and awareness. In order to compare and contrast the two, I discuss both together with regards to the open-ended responses and scale ratings.

6.4.2.1 Open-ended Responses

In response to questions about /ð/-stopping and /oʊ/-backing, most of the sample reported being aware of these (or similar) pronunciation variants in *some* Asian Americans' speech. Out of the 60 respondents, forty-two (70.0%) thought that "words like *that*, *though*, *brother*, and *rather* (with a "th" sound)" were pronounced differently by Asian Americans than white Americans. Forty (66.7%), responded the same for "words like *no*, *grow*, *home*, and *boat* (with an 'o' sound)".

When prompted to explain what they thought the difference was for those /ð/ words, 27 of the 42 (63.3%) specifically mentioned the "th" sound being pronounced like "d" (excluding cases where "th" was described as sounding like "t"). This indicates that /ð/-stopping was often available for metalinguistic discussion and quite accurately identified.

For /oʊ/ words, 29 out of 40 (72.5%) offered a specific explanation for how the pronunciation is different for Asian Americans, but these included a variety of descriptors that sometimes appear to conflict with those mentioned by others. For example, the most common description was that /oʊ/ sounded more emphasized (n = 11; including similar descriptors like "prominent" or "accentuated") but some also described /oʊ/ as *less* emphasized (n = 4). Indeed, these impressions may even coexist as one respondent mentioned that they believed "some Asian Americans [pronounced the 'o' sound] either with less or more emphasis than white Americans do". The next most common response described /oʊ/ as elongated (n = 8), but again there were responses that appeared to reference a shorter sound (n = 2). Additionally, some comments appeared to allude to vowel quality differences (n = 7) and others asserted that /oʊ/ sounded "rounder" (n = 2), "deeper" or "heavier" (n = 2), and "smoother" (n = 2) in Asian American speech.

This variety of responses may result from participants attending to different aspects of the vowel, or they may simply be using different words to convey the same impression. Either way, while these comments are difficult to translate into phonetic detail, at least some of these could be plausibly interpreted as relating to /oʊ/-backing, in particular impressions of more emphasis and a rounder or deeper sound. Overall, the idea of /oʊ/ being pronounced differently seems to be available for metalinguistic commentary to some extent of detail, but accuracy of the observations is an open question.

6.4.2.2 Scale Ratings

To more directly assess variation in awareness of /ð/-stopping and /oʊ/-backing, we can consider the scalar ratings that participants provided after being given a specific description the variants (for wording of the descriptions, see Appendix D). Just like the responses for the general idea of an Asian American accent, all awareness-related questions elicited a wide range of responses.

Based on composite AWARENESS scores, /ð/-stopping was on average moderately accessible to participants (*Median* = 4.2, M = 3.88, SD = 1.66), with variation found across individuals (range: 1–6.6). One general pattern could be seen in the individual scales: While a majority of participants reported being quite aware of /ð/-stopping in Asian American speech, there was a consistent minority that reported not being aware of it at all. For example, the FREQUENCY at which the concept of /ð/-stopping in Asian American speech came up in daily experiences varied a lot across individuals. Specifically, while a large proportion of the sample used the middle of the scale—responding with 4 (*sometimes*; 20.0%) or 5 (*often*; 18.3%)—a smaller but still sizable group selected either 1 (*never*; 16.7%) or 2 (*rarely*; 13.3%). A similar, if more exaggerated, pattern also was observed for FAMILIARITY with /ð/-stopping, which was relatively high on average. Many participants reported high familiarity on the upper end of the scale, most commonly selecting 7 (*extremely familiar*; 23.3%), 6 (*very familiar*; 21.7%), and 5 (*quite familiar*; 16.7%). However, a constant proportion reported no FAMILIARITY, selecting 1 (*not at all familiar*; 16.7%). In terms of noticing examples of /ð/-stopping in their daily life, respondents generally used the lower half of the scales, especially for MEDIAEXAMPLES.

In comparison, the idea of /oʊ/-backing as a feature of Asian American speech was only somewhat in AWARENESS based on the composite score (*Median* = 3.2, *M* = 3.29, *SD* = 1.67), lower on average than /ð/-stopping and with a slightly lower ceiling value (range: 1–6.2). From the individual scales, we can observe a couple of differences that support the interpretation that, despite a similar range of variation and some respondents reporting high awareness, /oʊ/-backing is a variable with less awareness attached on the whole. First, FREQUENCY of encountering /oʊ/-backing as a concept was much lower on average. Instead of selecting moderate options, responses about FREQUENCY were focused on the lowest end of the scale: Half of the respondents selected 1 (*never*; 25.0%) or 2 (*rarely*; 25.0%). Second, the average FAMILIARITY of /oʊ/-backing was also clearly lower. In contrast to ratings being concentrated in the upper range of the scale for /ð/-stopping, FAMILIARITY responses for /oʊ/-backing were either 1 (*not familiar at all*; 28.3%) or evenly distributed across range of options. Finally, while the distribution of the other scales had similar shapes between variables, the responses for /oʊ/-backing were generally more bottom heavy.

To supplement our understanding of these ideologies, I turn to the additional scales to identify potential attitudinal differences between the two variables. For ACCURACY of /ð/-stopping, most believed that some Asian Americans actually use /ð/-stopping in their speech (*Median* = 4.0, *M* = 4.28, *SD* = 1.66). Besides the most common response of 4 (*moderately true*; 21.7%), the top answers were on the upper half of the scale, including 7 (*extremely true*; 16.7%) and 6 (*very true*; 16.7%). Notably, however, /ð/-stopping was seen in a negative light by many participants (VALENCE; *Median* = 4.0, *M* = 3.48, *SD* = 0.75): While the most common response was 4 (*neutral*; 56.7%), all but one of the remaining responses indicated negative valence, including selections of 3 (*slightly negative*; 31.7%), 2 (*negative*; 8.3%), and 1 (*extremely negative*; 1.7%). Lastly, ALIGNMENT with /ð/-stopping was very low, potentially linked to the negative VALENCE. That is, this sample tended not to see themselves as using /ð/-stopping (*Median* = 1.0, *M* = 1.56, *SD* = 1.15), with the vast majority selecting 1 (*never*; 80.0%) for OTHERASCRIPTION as well as 1 (*not at all*; 73.3%) for SELFASCRIPTION.

Compared to /ð/-stopping, ideologies about /oʊ/-backing appear to be slightly different. First, participants were less likely to consider the idea of /oʊ/-backing to be very accurate (ACCURACY; *Median* = 4.0, *M* = 3.77, *SD* = 1.73), at least as it was described in this study. The most common responses for ACCURACY lay in the low-to-mid range of the scales, including top responses of 4 (*moderately true*; 23.3%), 2 (*slightly true*; 20.0%), and 5 (*quite true*; 18.3%). Second, VALENCE offered a clear difference between the two variables: Unlike ð/-stopping, /oʊ/-backing was considered largely neutral without any major negative or positive skew (*Median* = 4.0, *M* = 3.85, *SD* = 0.68). In terms of ALIGNMENT, the overall pattern for the two variables were similar, though there was slightly higher self-attribution for /oʊ/-backing compared to /ð/-stopping. That is, respondents generally did not think they used /oʊ/-backing in their speech (*Median* = 1.0, *M* = 1.55, *SD* = 0.96), but the proportion of the sample who responded 1 (*never*; 75.0%) to OTHERASCRIPTION and 1 (*not at all*; 65.0%) to SELFASCRIPTION was less extreme than for /ð/-stopping.

On the whole, experiences with ideologies about /ð/-stopping and /oʊ/-backing in Asian American speech appear to be highly variable across individuals, encompassing the full range of possibilities. For both variables, a portion of the sample was not at all aware of these ideologies, resulting in a somewhat bimodal distribution that represented two groups: people who were clearly aware and people who were not. Considering both the number of people who were aware and the strength of reported awareness where it existed, awareness was not only more common but also more prominent for /ð/-stopping than /oʊ/-backing. This difference is particularly noticeable when considering subjective familiarity alone. In addition to increased awareness, /ð/-stopping may also be stigmatized to some extent, unlike /oʊ/-backing.

6.5 Results: Perceptual Behavior

In this section, I present the behavioral results of the auditory 2AFC speaker identification task for the critical variables. The first part details the photo selection results while the second part details the awareness results based on the same data. In both, /ð/-stopping is discussed before /oʊ/-backing.

6.5.1 Photo Selections

6.5.1.1 /ð/-stopping

To begin, I examine persona photo selection behavior for the /ð/-stopping variable. Table 6.8 summarizes the empirical data for each PERSONAPAIRING via the sample means (and standard deviations) of PropEA for the Mainstream VARIANT (PropEA_M), PropEA for the Non-mainstream VARIANT (PropEA_N), and the difference between PropEA for the two VARIANT conditions (PropEA_{N-M}), calculated by subtracting by-participant PropEA_M from PropEA_N. Additionally, the table reports three values representing the Bayesian model posterior distributions of the difference between Non-mainstream and Mainstream VARIANT conditions per PERSONAPAIRING: the median of the posterior distribution ($\hat{\beta}$), the 95% credible interval (95% CrI), and the probability that the effect is positive ($P(\hat{\beta}> 0)$).

The data are depicted in Figure 6.3a as the proportion of Ethnic, Asian, or Eth-

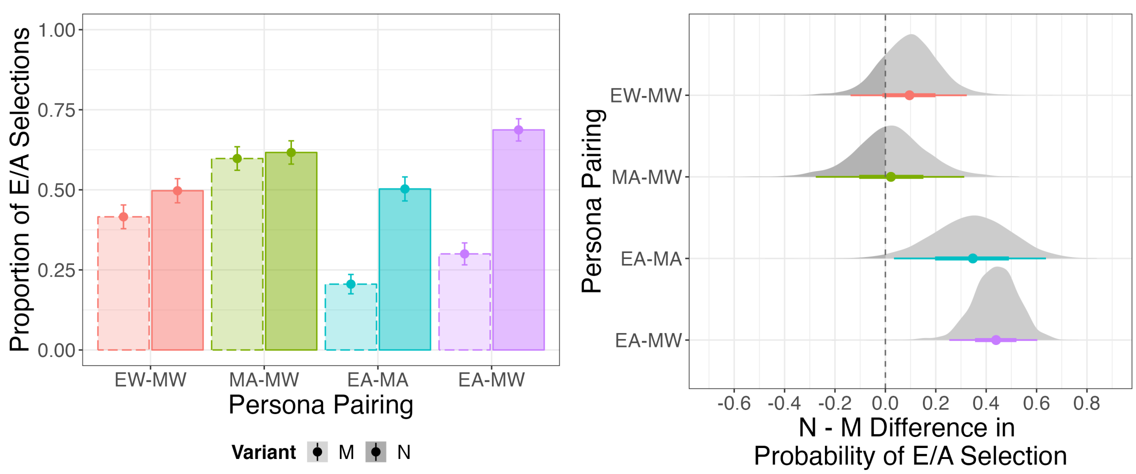

Figure 6.3: Group-level /ð/-stopping effects. (a) Mean proportion (error bars represent SE) of Ethnic, Asian, or Ethnic Asian photo selections (PropEA) per PERSONAPAIRING and VARIANT M = Mainstream, N = Non-mainstream) condition, based on the empirical data. (b) Bayesian posterior distributions of the difference in proportion of Ethnic, Asian, or Ethnic Asian photo selections between VARIANT conditions (N - M). Point intervals represent the median (point), 95% CrI (thin line), and 66% CrI (thick line). The proportion of the distribution below 0 is shaded with a darker grey.

nic Asian photo selections (y-axis) for each PERSONAPAIRING (x-axis) and VARIANT (shading) condition. In EA-MA and EA-MW PERSONAPAIRING contexts, the Non-mainstream variant condition elicited a greater proportion of target selections compared to the Mainstream variant condition. In contrast, we see no apparent difference in response patterns between VARIANT in the MA-MW PERSONAPAIRING context while there is a small VARIANT effect in the control condition.

The statistical model confirms these observations. Figure 6.3b presents the posterior probability distributions of the difference between the two VARIANT conditions (N-M) within each PERSONAPAIRING condition. There was strong evidence for a positive effect of VARIANT in both EA-MA and EA-MW PERSONAPAIRING conditions. Between these two, the effect size was larger when presented with EA-MW PERSONAPAIRING options. In other words, listeners were overall more likely to identify an Ethnic Asian photo as the speaker when they heard stopped /ð/ (e.g., *federweight*), a bias which was more pronounced when the other option differed in both culture and race/ethnicity (i.e., Mainstream White) rather than differing in culture only (i.e., Mainstream Asian).

No evidence of the predicted VARIANT effect was found for the MA-MW PERSONAPAIRING condition. As Figure 6.3b shows, the distribution is essentially centered on zero, suggesting that listeners were no more likely to select the Mainstream Asian photo over the Mainstream White photo when hearing stopped /ð/ than they were when hearing fricated /ð/. This lack of evidence is particularly striking given that a potential weak effect of VARIANT was observed in the control condition; this is interpreted as listeners selecting Ethnic White photos over Mainstream White photos slightly more often when hearing stopped /ð/. While this condition was not designed to be completely analogous to the target conditions, if this effect could be extrapolated with more certainty, it would suggest that some form of perceived Ethnic cultural alignment (e.g., being perceived as less

American) is the key social dimension associated with /ð/-stopping while Asian race/ethnicity alone is not.

Regardless, the main finding is clear: The presence of /ð/-stopping biased listeners to identify the speaker as someone who embodies an Ethnic Asian persona, representing AACs perceived as more oriented towards Asian culture. It did not result in a bias to identify the same speakers as Mainstream Asian. Thus, the general finding is that of a persona-specific effect rather than a race/ethnicity effect.

6.5.1.2 /oʊ/-backing

Next, I examine persona photo selection behavior for the /oʊ/-stopping variable. Table 6.9 summarizes the empirical data and model posteriors for each PERSONAPAIRING. Figure 6.4a shows that differences were observed in the proportion of Ethnic and/or Asian photos selected between Non-mainstream and Mainstream VARIANT conditions for the target PERSONAPAIRING contexts. While these were relatively narrow margins, they are in the predicted direction such that Non-mainstream variants elicited more Ethnic and/or Asian photo selections. Of the three, the MA-MW PERSONAPAIRING context elicited the smallest VARIANT difference. The EW-MW control condition, on the other hand, did not show any apparent difference in response patterns between VARIANT.

As illustrated by the posterior distributions in Figure 6.4b, the statistical model suggests that while the evidence is not especially strong, there is some evidence of a positive VARIANT difference (N-M) effect in the three target PERSONAPAIRING conditions; this trend is notably absent in the control condition. In particular, moderate evidence supports a modest VARIANT effect in the EA-MW condition, as the lower bound of the 95% credible interval is close to zero and the probability of a positive effect is over 0.95. Weak evidence is found for both MA-MW and EA-MA conditions, where the probability of a small positive effect is between 0.88 and 0.89.

Persona Pairing	**PropEA$_M$ (SD)**	**PropEA$_N$ (SD)**	**PropEA$_{N-M}$ (SD)**	$\hat{\beta}$	**95% CrI**	$P(\hat{\beta} > 0)$
EW-MW	0.417 (0.284)	0.503 (0.277)	0.086 (0.439)	0.096	[-0.138, 0.323]	0.804
MA-MW	0.600 (0.299)	0.617 (0.346)	0.017 (0.453)	0.022	[-0.276, 0.313]	0.565
EA-MA	0.206 (0.238)	0.503 (0.333)	0.297 (0.435)	0.347	[0.034, 0.637]	0.984
EA-MW	0.300 (0.272)	0.689 (0.324)	0.389 (0.460)	0.439	[0.254, 0.603]	1.000

Table 6.8: /ð/-stopping Photo Selection Summary Statistics per PERSONAPAIRING. To describe the empirical data: mean proportions *(SD)* of Ethnic, Asian, or EthnicAsian photo selections for the N VARIANT condition, the M VARIANT condition, and the by-participant N-M VARIANT difference. To describe Bayesian model output: median of the posterior distributions for the N-M VARIANT difference, 95% credible interval of the posterior distributions, and proportion of the the posterior distribution above zero representing the probability of a positive effect.

Persona Pairing	**PropEA$_M$ (SD)**	**PropEA$_N$ (SD)**	**PropEA$_{N-M}$ (SD)**	$\hat{\beta}$	**95% CrI**	$P(\hat{\beta} > 0)$
EW-MW	0.514 (0.327)	0.531 (0.275)	0.017 (0.428)	0.001	[-0.236, 0.249]	0.504
MA-MW	0.589 (0.327)	0.678 (0.294)	0.089 (0.397)	0.115	[-0.094, 0.342]	0.887
EA-MA	0.328 (0.278)	0.453 (0.293)	0.125 (0.380)	0.147	[-0.108, 0.448]	0.884
EA-MW	0.467 (0.326)	0.589 (0.338)	0.122 (0.402)	0.200	[-0.035, 0.428]	0.960

Table 6.9: /oʊ/-backing Photo Selection Summary Statistics per PERSONAPAIRING. To describe the empirical data: mean proportions *(SD)* of Ethnic, Asian, or EthnicAsian photo selections for the N VARIANT condition, the M VARIANT condition, and the by-participant N-M VARIANT difference. To describe Bayesian model output: median of the posterior distributions for the N-M VARIANT difference, 95% credible interval of the posterior distributions, and proportion of the the posterior distribution above zero representing the probability of a positive effect.

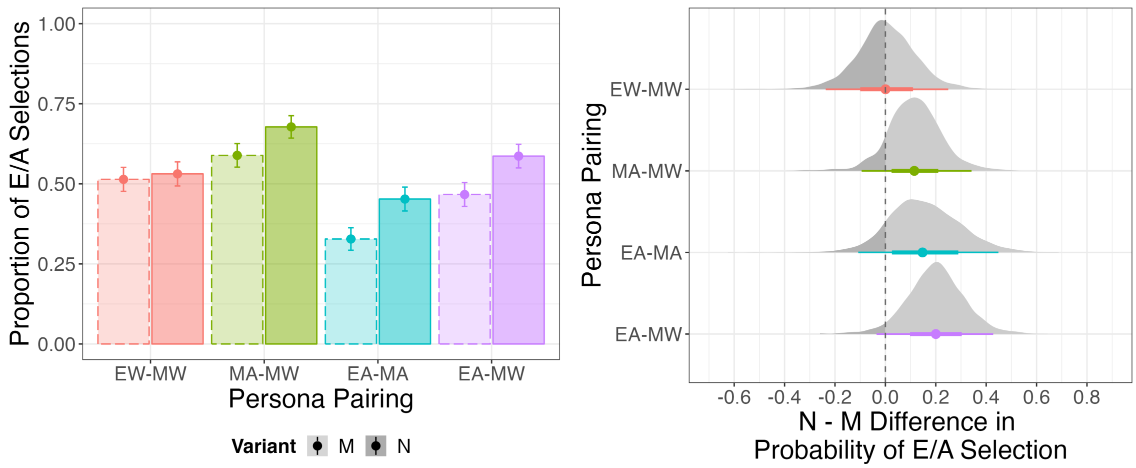

Figure 6.4: Group-level /oʊ/-backing effects. (a) Mean proportion (error bars represent SE) of Ethnic, Asian, or Ethnic Asian photo selections (PropEA) per PERSONAPAIRING and VARIANT M = Mainstream, N = Non-mainstream) condition, based on the empirical data. (b) Bayesian posterior distributions of the difference in proportion of Ethnic, Asian, or Ethnic Asian photo selections between VARIANT conditions (N - M). Point intervals represent the median (point), 95% CrI (thin line), and 66% CrI (thick line). The proportion of the distribution below 0 is shaded with a darker grey.

Although estimates of a VARIANT effect in the EA-MA condition are comparable to the other two target conditions, there is greater uncertainty in this particular PERSONAPAIRING context.

Taken together, these results suggest that listeners were slightly more likely to identify a Mainstream Asian or Ethnic Asian photo as the speaker when they heard backed /oʊ/. This subtle bias was more clearly seen when listeners selected between photos that differed on both the race/ethnicity and cultural dimensions (i.e., Ethnic Asian as opposed to Mainstream White). There was also potential evidence of this bias when listeners selected between photos that varied in perceived race/ethnicity alone (i.e., Asian vs. White) or cultural alignment alone (i.e., Ethnic vs. Mainstream). In these cases, the Asian or Ethnic photos were more commonly selected over the White or Mainstream photos when hearing backed /oʊ/. As a point of comparison, no evidence of a VARIANT effect was found for the control EW-MW PERSONAPAIRING condition: Listeners were not systematically selecting one White photo over the other when hearing backed /oʊ/ compared to fronted /oʊ/.

To summarize, the presence of /oʊ/-backing seemed to bias listeners to identify a speaker as Asian, especially when choosing between Ethnic Asian and Mainstream White personae. A similar bias may lead listeners to identify the speaker as a more ethnically-oriented Asian individual compared to a more mainstream-oriented Asian individual. More evidence is needed to support these findings. Generally, these results point towards /oʊ/-backing as a phonetic feature linked to both race/ethnicity-related and culture-related effects in the AAC context.

6.5.2 Awareness

6.5.2.1 Variable-level

Moving on to the awareness results, the first approach I present is qualitatively comparing results for the two sociophonetic variables that differ in awareness. Given that self-reported awareness for /ð/-stopping was on average greater than awareness for /oʊ/-backing, this predicts larger sizes of the target effect in the /ð/-stopping condition.

Effects sizes of /ð/-stopping compared to /oʊ/-backing can be assessed by judging the difference in proportion of Ethnic and/or Asian photo selection between VARIANT (N-M) conditions (Table 6.8 and 6.9). The prediction appears to be borne out. Since the two variables had a slightly different pattern in results, I only consider the two PERSONAPAIRING conditions that show the VARIANT difference effect for both variables. These were the two conditions involving the Ethnic Asian persona. Comparing across variables in the EA-MW PERSONAPAIRING condition, the mean difference in proportions between Non-mainstream and Mainstream VARIANT conditions is nearly 0.4 for /ð/-stopping (M = 0.389) while the mean difference for /oʊ/-backing is around a third of that size (M = 0.122).

Similarly, in the EA-MA condition, the difference in proportions is greater for /ð/-stopping (M = 0.297) compared to /oʊ/-backing (M = 0.125).

I interpret these findings as such: When participants heard the speaker produce a variable that has stronger awareness attached, in this case /ð/-stopping, they showed a more pronounced bias to identify the speaker as Ethnic Asian. When they heard a variable with weaker awareness attached, in this case /oʊ/-backing, they seemed to show the same bias but to a lesser extent. Generally, this is consistent with the possibility that strength of awareness mediates strength of sociophonetic associations.

6.5.2.2 Individual-level

The second approach I take is examining the relationship between individual-level awareness and perceptual behavior. The slope estimates, 95% credible intervals, and probability of a positive effect per PERSONAPAIRING for both /ð/-stopping (Table 6.10) and /oʊ/-backing (Table 6.11) are provided. Figure 6.5 depicts individual data points representing each participants' mean VARIANT proportion differences (y-axis) and their level of awareness (x-axis); this is overlaid with posterior prediction slopes based on the statistical models.

Based on visual inspection of the data, we observe some numerical trends of a positive relationship between awareness and response bias for four of the target conditions: EA-MA and EA-MW for /ð/-stopping and MA-MW and EA-MW for /oʊ/-backing. This trend, if supported by enough evidence, would be consistent with the following interpretation: As an individual's degree of awareness for either /ð/-stopping or /oʊ/-backing increases, the likelihood that they tended to select the Ethnic and/or Asian photo specifically when hearing the Non-mainstream VARIANT also increases. Notably, although there was weak evidence for a photo selection bias effect in the EA-MA condition for /oʊ/-backing, there seems to be no trend of an individual-level awareness correlation. Regardless, when the model output is considered, we find that the uncertainty is very high for all conditions and thus the evidence for a positive correlation is not particularly compelling.

In other words, these data show no clear evidence for a relationship between higher awareness of a particular sociophonetic variable—either /ð/-stopping or ʊ/-backing—and the predicted photo selection bias for the (Ethnic) Asian photo when hearing that variable. However, there appears to be a consistent but small trend in the predicted positive direction.

Overall, these results provide a step in the right direction, although the role of awareness are still inconclusive. Based on qualitative observations, individuals who have higher awareness of /ð/-stopping or /oʊ/-backing *may* be associated

Persona Pairing	$\hat{\beta}$	**95% CrI**	$P(\hat{\beta} > 0)$
EW-MW	-0.008	[-0.236, 0.211]	0.474
MA-MW	-0.029	[-0.260, 0.217]	0.396
EA-MA	0.133	[-0.177, 0.400]	0.832
EA-MW	0.095	[-0.279, 0.453]	0.705

Table 6.10: /ð/-stopping awareness summary statistics per PERSONAPAIRING. Median of the posterior distributions for the interaction between awareness and N-M VARIANT difference, 95% credible interval of the posterior distributions, and proportion of the the posterior distribution above zero representing the probability of a positive effect.

Persona Pairing	$\hat{\beta}$	**95% CrI**	$P(\hat{\beta} > 0)$
EW-MW	0.020	[-0.177, 0.212]	0.586
MA-MW	0.045	[-0.166, 0.253]	0.677
EA-MA	-0.003	[-0.224, 0.223]	0.488
EA-MW	0.112	[-0.153, 0.384]	0.803

Table 6.11: /oʊ/-backing awareness summary statistics per PERSONAPAIRING. Median of the posterior distributions for the interaction between awareness and N-M VARIANT difference, 95% credible interval of the posterior distributions, and proportion of the the posterior distribution above zero representing the probability of a positive effect.

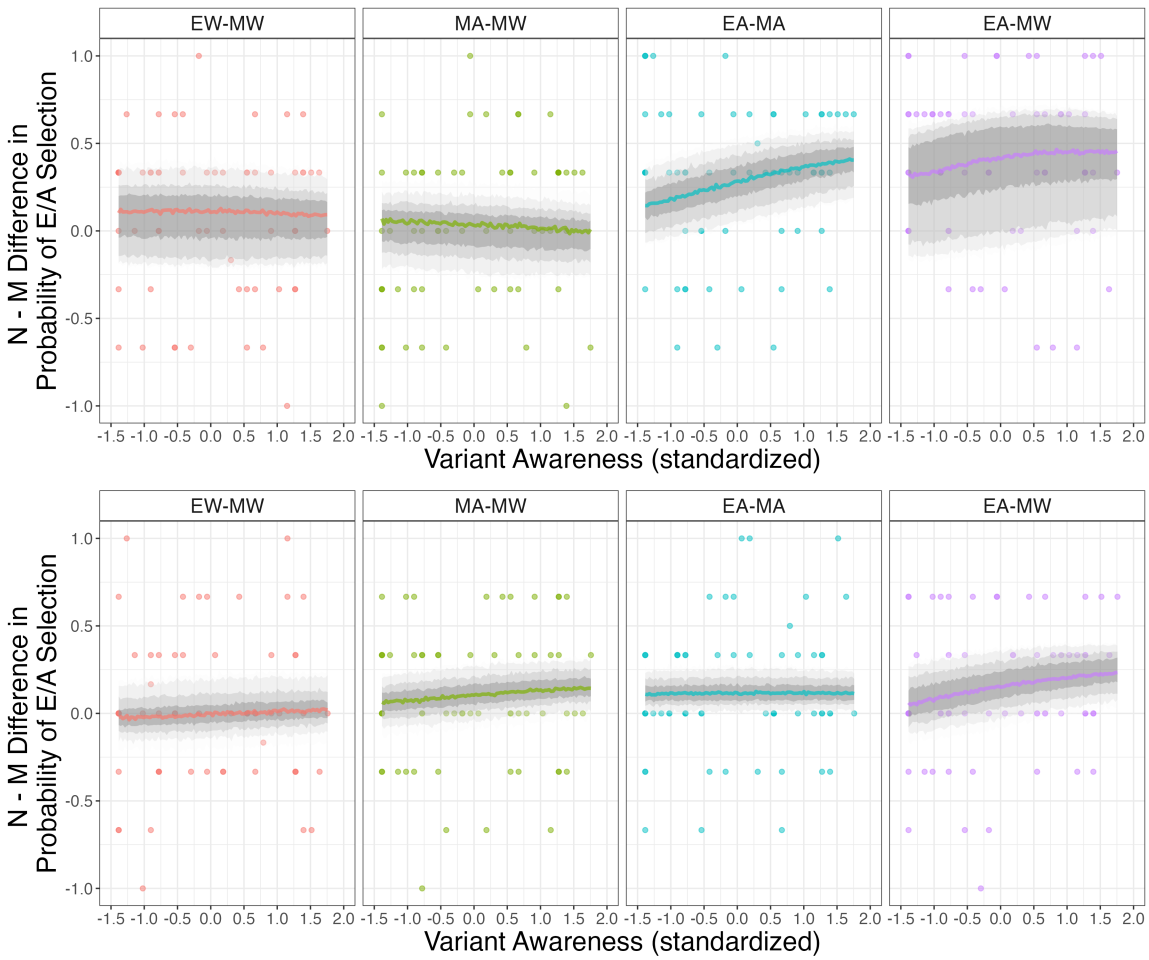

Figure 6.5: Individual-level awareness effects per PERSONAPAIRING for (top) /ð/-stopping and (bottom) /oʊ/-backing. Posterior predictions of the effect of awareness on the difference in probability of Ethnic, Asian, or Ethnic Asian photo selections between N and M Variants (N - M), based on the Bayesian regression model. Ribbons represent 95% CrI (light grey), 80% CrI (medium grey), and 50% CrI (dark grey). Points represent by-participant N-M differences in PropEA per Persona Pairing, based on the empirical data.

with a larger bias to identify the speaker as (Ethnic) Asian when hearing the variable. More data would need to be collected to reduce the uncertainty and confirm whether these patterns hold out in a larger sample.

6.6 Discussion

This study directly followed up on Chapter IV with more controlled stimuli and methods, directly assessing ideologies and awareness of /oʊ/-backing and /ð/-stopping and testing the behavioral effect of these variants as perceptual cues to certain types of Asian American speaker identity. At the same time, I tested hypotheses about the role of personae and awareness in the encoding of sociophonetic representations by examining their interaction with perceptual speaker identification behavior. In general, evidence was found in support of /oʊ/-backing and /ð/-stopping as ideologically and behaviorally linked to AAC identities. In addition, some evidence was found to be consistent with the theorized role of personae and awareness in mediating sociophonetic representations and perception, though in other cases there was too much uncertainty to come to a conclusion. In the following, I first discuss what the findings tell us about Asian American speech, including the differences that we see between the two target variables. Then, I discuss the implications of the tentative findings about personae and awareness for sociophonetic cognition.

6.6.1 Phonetic Features and Ideologies of AAC Speech

These results offer novel empirical evidence that specific linguistic variables may be socially indexed to specific Asian American personae. To the best of my knowledge, this is the first study that provides controlled experimental evidence from perceptual behavior to identify specific phonetic cues contributing to an "Asian American accent" or "sounding Asian". Moreover, it is the first study to explore further the degree of awareness for specific variables and the connections to specific subcategories of Asian Americans, both of which are important to fully understanding what the phenomenon of "sounding Asian" is.

For Asian Americans from California, /oʊ/-backing appears to be a perceptual cue indexed to Asian or Asian American identity. When hearing backed /oʊ/ in a word, they were more likely to guess that the speaker was Asian (compared to white), especially one who looked more ethnically-oriented. It is worth mentioning here that the stimuli largely consisted of East Asian photos and voices, so this result could potentially be specific to East Asian Americans. Regardless, many respondents were also able to identify and discuss /oʊ/-backing as a feature of an Asian American accent in metalinguistic commentary. This corroborates previously reported metalinguistic commentary results where /oʊ/ is occasionally mentioned as being pronounced differently by Asian Americans (e.g., Bauman, 2016; A. Cheng & Cho, 2021), and further shows that awareness of /oʊ/-backing may be more widespread than previous reports would suggest. This is likely because the current study asked directly about /oʊ/ rather than only asking open-ended questions about features of Asian American speech, where less specific or more socially salient features would take the main focus. Here, I found that a sizable proportion of Asian Americans both believe in and can describe phonetic variation of /oʊ/ with regards to Asian American speech.

Additionally, even though degree of awareness was quite variable, there were some respondents for which /oʊ/-backing seemed to be very consciously accessible, suggesting that it is a socially salient feature to some people. In A. Cheng and Cho (2021), a Korean American respondent commented that /oʊ/ sounded "rounder" for Asian Americans. Interestingly, Korean Americans were the only specific ethnic group mentioned in ideological descriptions of /oʊ/ in the current study: One participant posited that Korean Americans used a "very pronounced" /oʊ/ "with a deeper, longer vowel sound", while another thought that Korean Americans pronounced /oʊ/ "with a roundness and vigor that is distinct". Assuming that these ideologies are sourced at least in part from direct exposure to

Asian American speakers, the metalinguistic and behavioral results are also consistent with studies that report /oʊ/-backing in Asian American speech production (e.g., Bauman, 2016), particularly documented in Korean Americans (e.g,. A. Cheng et al., 2016) with more ties to ethnic identity (e.g., L. Jeon, 2017; A. Cheng, 2020; D.-E. Kim, 2021).

On the other hand, /ð/-stopping seems to perceptually cue Asian Americans who are specifically more oriented towards their ethnic culture, such as those who may be perceived as a "FOB". The association between /ð/-stopping and at least some AAC identities is consistent with previous reports of metalinguistic commentary about Asian American voices (A. Cheng & Cho, 2021) as well as post-hoc acoustic analyses of speech samples found to often be identified as Asian (P. Wong & Babel, 2017; A. Cheng & Cho, 2021). Like open-ended comments about /oʊ/, only one specific ethnic group was mentioned specifically. In this case, Vietnamese Americans were singled out as having a link to /ð/-stopping by two respondents. To the best of my knowledge, no production studies have examined /ð/-stopping in Asian American speech, and few have examined Vietnamese American speech practices in general (for one example, see Nguyen, 2019). This may be an interesting thread for future research to follow up on.

Because this study only found positive evidence of a sociophonetic association between /ð/-stopping and an Ethnic Asian persona, this could suggest that /ð/-stopping is not at all ideologically associated with more mainstream-oriented Asian identities. Note that this would not mean that "mainstream" AACs never produce /ð/-stopping, only that it is not represented as part of ideologies about mainstream-oriented Asian American speech. If we also assume that someone with a mainstream-oriented white persona would not be expected to produce /ð/-stopping, this suggests that the lack of Asian photo selection bias when choosing between Mainstream Asian and Mainstream White personae was due to participants making a final decision randomly, as neither option seemed correct to listeners. However, I would be cautious in making this conclusion before ruling out some potential alternative explanations for these results. I discuss here three possibilities that could be further explored in future research to better understand the /ð/-stopping as an AAC sociophonetic variable: (i) stigma and response strategies, (ii) phonetic details of the auditory stimuli, and (iii) contextual effects of speaker voice and other intersectional factors.

First, /ð/-stopping was specifically rated by many Asian Americans as being a negative feature of speech. Moreover, when asked about their thoughts on the target Asian American personae, participants largely self-identified as "normal" or "bicultural" Asian American, which was represented by the Mainstream Asian photos in the experiment. In this case, one possible explanation is that some listeners chose not to select the Mainstream Asian photo when hearing stopped /ð/, even when their first instinct was the Asian photo, because they did not want to identify their own in-group as using a stigmatized feature. That is, this explanation assumes that /ð/-stopping *is* socially indexed to Mainstream Asian, but because of negative attitudes towards the feature, participants (consciously) suppressed their initial responses which obscured the associative link in behavior. If this were the case, a different behavioral task that assesses social association with more automaticity and less explicit social categorization may show a different result.

A second possibility is that this result is a reflection of the particular phonetic profiles of the /ð/-stopping stimuli heard in the experiment. The current stimuli generally represented articulations that sounded quite clearly flapped, based on auditory judgments. It could be that such an extreme deviation from a mainstream fricative /ð/ variant is not associated with a Mainstream Asian character but a more intermediate realization of /ð/-stopping might be. Similarly, it could be that intervocalic /ð/-stopping is less associated with Mainstream Asian than

/ð/-stopping in another position would be, such as word-initial onset /ð/ (e.g., in words like *this*, *those*, and *there*). If so, a similar task could be conducted with different conditions for various realizations of /ð/-stopping (e.g., stronger or weaker articulations) to investigate whether a different or more subtle version would be associated with Asian Americans who are generally perceived as aligned with the American mainstream.

A third potential reason that /ð/-stopping was not found to be associated with mainstream-oriented Asian Americans could be linked to interactions with the voices and photos used as stimuli. For example, each of the target voices elicited somewhat different social impressions, including some of the voices being perceived as most likely white American while others were mostly perceived as Asian American based both on norming and overall response patterns per speaker. In this case, it is conceivable that /ð/-stopping might be able to elicit a more consistent Mainstream Asian selection bias in the context of certain voices but not others.[15] Relatedly, it could be that /ð/-stopping is associated with *some* Asian Americans who would be considered mainstream-oriented, but not the ones represented in this study. For instance, most of the speakers identified as East Asian, most of the photos were also interpreted as East Asian, and there were only a few photos per persona. If this variable was associated more with, say, Southeast Asian ethnic groups like Vietnamese Americans, the context of the voices and photos may have been a relevant factor. In general, the presence of /ð/-stopping might play a role in influencing behavioral responses only if other factors were also present such that the combination of features together give rise to a Mainstream Asian persona impression, or to more specific personae not captured in this study. Future research can examine /ð/-stopping ideologies in more detail to uncover the ways that it does or does not link to various AAC identities, including complementary work

[15]There was indeed variation across speakers, but there was also a variety of other differences between them so the effect of speaker voice cannot be isolated.

examining the extent to which /ð/-stopping may occur in Asian American speech.

6.6.1.1 Limitations for Understanding AAC Speech

Next, I discuss a few other considerations regarding limitations of the current study on the topic of AAC speech, along with thoughts about future directions. One barrier to generalizing these results is the particular demographic sampled for inclusion. Californian Asian Americans were selected specifically because of a higher concentration of co-ethnics in the region, increasing likely contact with other Asian Americans. This may make them particularly unique compared to most other Asian Americans, especially with regards to degree of awareness.

The sample was also not particularly balanced in terms of Asian ethnic background, as any who self-identified as East or Southeast Asian could participate. A large proportion of the sample came from a small set of ethnic groups. The results that we find can be expected to differ based on individual experiences with what would fall under the label of "Asian American speech", especially if they have varied amounts of experience with specific ethnic varieties. For example, if Korean Americans do indeed produce more /oʊ/-backing and it is a more prominent ideological association for this group in particular, then we would expect listeners with different amounts of experience with Korean American speech to also differ in their ideologies and response patterns. Because this study probes ideologies about "Asian American" speech defined generally, participants were likely drawing on different types of Asian American exemplars that were not necessarily consistent across the sample. This likelihood is supported by some of the ideology comments highlighting specific ethnic backgrounds, as well as direct mention of variation in phonetic realizations depending on "what kind of Asian American is saying it". That the results showed group-level effects, particularly apparent for /ð/-stopping, suggests that there is or can be some generalizability or consistency in ideologies across Californian Asian Americans with different ethnic backgrounds and experiences. It would be interesting to explore whether a more diversified Asian American sample, both in location and ethnicity, would show similar outcomes in ideology and behavior, and whether a systematic comparison of different ethnicities would reveal differences between what they think is an "Asian American accent".

In a similar vein, we would expect that experiences with and ideologies of Asian American speech are likely to differ between Asian Americans and other Americans on the whole (e.g., A. Cheng & Cho, 2021), due to both typical amount of exposure as well as other attitudinal factors. Thus, how /oʊ/-backing and /ð/-stopping may act as perceptual cues to listeners of different social backgrounds remains to be further investigated, including the extent that each of these may be in awareness and the manner in which these associations may have been acquired. A better understanding of these ideologies for both Asian Americans and others, as well as how it impacts perceptual behaviors, would have important implications for how AACs experience social interactions, including with regards to linguistic profiling and discrimination (e.g., Craft et al., 2020).

With regards to the stimuli, another caveat—like that of the East Asian bias in voices and photos—is the inclusion of only voices and faces perceived as women. This limits our generalizable conclusions about "sounding Asian" as there can be intersectional differences, with gender being one dimension that has been related to differences in production and perception: Some Asian American production studies have noticed a pattern of women producing more /oʊ/-backing (A. Cheng, 2020; D'Onofrio & van Hofwegen, 2020), and there are also gender-linked ideological associations like higher pitch associated with Asian identity for women in particular (A. Cheng & Cho, 2021). Other social identities that intersect with Asian American identities could also be relevant, so future work would benefit from including and exploring these different ways of "being Asian" and how that connects to the idea of "sounding Asian".

In terms of /oʊ/-backing as represented in the auditory stimuli, one key element of the design was natural productions by each of the speakers rather than acoustic manipulation. While artificial manipulation could have led to more control over non-target differences between Mainstream and Non-mainstream variant conditions, it was less feasible due to the large number of speakers and tokens included in the design. This meant that while F2 crucially differed between conditions, and some other features were assessed to be relatively matched (e.g., pitch contour), there may have been other differences at play. In particular, I did not specifically control for /oʊ/-monophthongization. While this variable was not found to be particularly consistent across the production of Asian American speakers in Chapter III, previous studies have reported production results indicating that it could be associated with (East) Asian Americans (e.g., D'Onofrio & van Hofwegen, 2020; D.-E. Kim, 2021) . Also, in ideological commentary about how Asian Americans pronounce /oʊ/, some participants did provide descriptions that could be reasonably interpreted as monophthongization, such as /oʊ/ sounding more "one note". Some amount of /oʊ/-monophthongization could have been present in the Non-mainstream variants along with /oʊ/-backing, and it could potentially play a role in inducing an impression of Asian American identity, either alone or together with /oʊ/-backing. This study does not tease these possibilities apart, but future studies could work to disentangle these articulatory elements to be more precise about the acoustic-perceptual details that are sociophonetically associated with Asian American identities.

Finally, it is worth noting that the sociolinguistic questionnaire was ordered after the experimental task where participants heard many examples of these variants paired with Asian and white faces; this may have resulted in inflated rates of self-reported conscious access to the ideologies due to contextual salience, compared to if the questionnaire was not preceded by the task. However, this would likely only happen if there was some existing awareness in the first place. Thus, any effect of task ordering could also be viewed as simply boosting signals to ensure that awareness is detected more accurately, especially since most of the questions ask respondents to recall and enumerate past experiences rather than providing in-the-moment subjective ratings. In addition, all participants received the same amount of exposure in the task (i.e., number of trials per condition) so we might still expect variation in awareness across individuals to be preserved. Regardless, future work examining ideology and awareness of these variables may wish to use different methods to avoid this ambiguity.

6.6.2 Personae as Social Information

Turning to the role of personae in sociophonetic perceptual behavior, the results of this study highlight how microsocial persona-based information can be tied to mental representations of phonetic detail. A key finding was that listeners behaved differently in response to the same auditory stimuli involving /ð/-stopping when paired with photos of Asian faces portraying different visual styles: They were more likely to select a photo representing an Ethnic Asian persona than a Mainstream Asian persona when they heard a word with stopped /ð/. This indicates that, while processing speech, listeners are activating mental associations between specific linguistic information like /ð/-stopping in the auditory domain and specific social information about a persona or type of person based on visual elements like perceived race/ethnicity, clothing, hairstyle, and makeup. Theoretically, this implies that listeners have different mental models for the speech of what can be rather narrowly-defined microsocial categories of people (e.g., Ethnic Asian), in between the specificity of individual speakers and the abstraction of broad macrosocial dimensions. Further, this may underlie (or be an alternative to) processing characterized as relying on macrosocial categories (e.g., Asian).

The results for /ð/-stopping corroborate previous studies suggesting that social personae are ideological constructs that listeners use to make sense of sociophonetic variation in perception (e.g., MacFarlane & Stuart-Smith, 2012; D'Onofrio, 2018b; Leigh, 2021). In this study, I found support for a similar persona-based effect using a different task and design, namely a speaker identification task with multiple voices and multiple photos. The fact that this effect was evidenced across various methods and sociolinguistic contexts, and especially in the current study where social information was conveyed via photos of people one might realistically encounter, strengthens the proposal that listeners use this type of information in real-world social interactions.

Whether social processing is always mediated by persona-level information is still an open question, as is the question of which personae are in play. For example, the presence of /oʊ/-backing led to largely similar behavioral patterns when provided with either or both Ethnic Asian and Mainstream Asian response options. This could mean that this phonetic variant is associated with an abstract pan-ethnic "Asian American" category or macrosocial "Asian" category at the representational level. A second possibility is one where /oʊ/-backing is tied to persona-based representations for a stereotyped "Asian" and/or "Asian American" category (similar to stereotyped "New Yorker" or "Chicagoan" personae, as in Becker, 2014; D'Onofrio, 2021a). Another option may be that /oʊ/-backing sociophonetic associations are represented with links to both of the narrower Asian American personae presented in this study (as well as any others that exist in their ideologies). Could these forms of social information tagged to phonetic detail all be available to listeners at the same time, or might it only be personae-based representations that are available as social information, even if these personae represent multiple levels of abstraction? If so, how are different levels of representation called upon at different times? Future research could seek to delve deeper into understanding ideologies of different types of people within social groups, including both personae representing both broader and narrower instantiations within a macrosocial category, then test to what extent those social types are able to explain patterns of sociophonetic perceptual behavior.

Another way to frame the results is that listeners responded to the Ethnic Asian persona differently from Mainstream Asian, while the Mainstream Asian persona was treated similarly to Mainstream White. This is reminiscent of the pattern of results found in D'Onofrio's (2019) primed lexical recall task, where listeners performed similarly when given a more American-perceived Asian photo and a White American photo while they showed a different behavior when primed with a more foreign-perceived Asian photo. Methodologically, this calls for caution in the interpretation and generalization of experimental research that manipulates decontextualized race/ethnicity, such as studies that compare behavioral differences based on exposure to an (East) Asian face and a white face where "Asian" is expected to be associated with a "non-native" English accent (e.g., Rubin, 1992, McGowan, 2015; see discussion in D'Onofrio, 2019 and Chapter V). The current results provide another example demonstrating that listeners do not bring the same linguistic expectations to a listening task for all people who are racialized as Asian, but instead can tap into expectations based on aspects like social context and stylistic presentation.

Because of this, using stimuli that are intended to "isolate" Asian race/ethnicity may simply lead to more ambiguity. At best, this practice may elicit heterogenous interpretations of the "same" macrosocial category, or of a persona representing a stereotypical version of that macrosocial category. At worst, it may elicit heterogenous interpretations of "different" microsocial categories housed within the macrosocial category, such as a stereotypical "Asian", Ethnic Asian, or Mainstream Asian, that vary with the experience and context of the listener. Either way, we may wish to be careful about making generalizations about what people associate with certain social groups in different contexts, and be skeptical of how those results may translate to social interactions in real world contexts. These findings once again highlight the importance of considering the influence of personae and other conceptions of microsocial or intersectional information when theorizing about sociophonetic processing or speech perception—particularly if the goal is to ultimately understand how people process speech in their daily lives.

6.6.3 Awareness in Sociophonetic Perception

Finally, this study offers new evidence using direct and gradient measures that degree of awareness may be one factor linked to the strength of sociophonetic associations, as predicted by attention-based exemplar approaches. At the group level, /ð/-stopping elicited clearer and larger effects of hearing the variant on selecting Ethnic Asian persona photos than /oʊ/-backing did. This fully aligns with predictions based on comparatively higher levels of awareness for /ð/-stopping across participants on the whole. At the individual level, results were less clear due to high uncertainty, but all the conditions that showed trends were in the predicted positive direction, and none were in the negative direction. At the very least, this provides some new information that can be used to build off of in future research on this topic.

These results are consistent with the findings of Juskan (2018), where higher salience sociolinguistic variables in Standard Southern British English resulted in predicted social priming effects but lower salience variables did not; the author assessed salience for each variable based on a holistic assessment of production patterns and metalinguistic commentary. In a previous study, I used a questionnaire method similar to the current study to more directly measure awareness in a gradient fashion for one variable, /aʊ/-raising in Michigan (L. S. P. Cheng, 2021). The current results are also generally consistent with that study, which found some evidence for an individual-level effect of measured awareness on social priming outcomes, conditional on another individual-level factor.

In the current study, I combined these approaches to directly measure gradient awareness for multiple variables using the same methods, and was able to show that a variable with higher awareness elicited larger speaker identification bias effects while a variable with lower awareness elicited the same effects at a smaller scale. In other words, my results add to existing literature by providing more nuanced evidence suggesting that gradience in awareness relates to gradience in sociophonetic behavior, as would be expected. This work moves the literature beyond inferring salience or awareness based only on past descriptions of community-level salience and high-level metacommentary, as well as only taking categorical approaches (e.g., Lawrence, 2015, Chang, 2017). Those approaches do not provide enough precision to make informed judgments about what the results tell us, as sociocognitive factors like awareness and salience vary continuously across individuals based on specific experiences and should also influence sociophonetic represention and processing in a continuous manner. As the ideology results demonstrate, even if a larger proportion of people report that they are aware of a sociolinguistic variable like /ð/-stopping and /oʊ/-backing, the degree to which the variable is accessible to them is highly variable and distributed continuously. This must be accounted for in our treatment of sociophonetic cognition.

Overall, these results fall in line with predictions implied by Drager and Kirtley's (2016) attention-weighting exemplar model, where awareness is one factor that may increase salience and attention, which in turn increases the strength of encoding and activation of sociophonetic associations. Under this approach, I would interpret these results as follows. Based on past experiences that were not explored in this study (e.g., interactional exposure; previous experiences with metalinguistic commentary or performance), the association between /ð/-stopping and Asian American identities was tied to more conscious access, or awareness, for many of the Asian American individuals in this study. That would mean that /ð/-stopping exemplars linked to ethnically-oriented Asian American speakers were stored in memory with higher weights compared to /oʊ/-backing exemplars linked to ethnically-oriented Asian American speakers. During the speaker identification task, we might then expect that /ð/-stopping would readily activate associations with a Ethnic Asian persona, leading to high rates of Ethnic Asian photo selections. On the other hand, /oʊ/-backing would more weakly activate the same associations, lead to comparatively lower rates of Ethnic Asian photo selections, though still occurring somewhat more often than not.

Next, I discuss some limitations and future directions. This study was not clearly able to detect the effect of awareness in some cases, especially at the individual level. In terms of the awareness measurement instrument, it could be that the questions probing the construct of awareness were not ideal, and that other questions could have done a better job of targeting the construct. Another limitation is that I did not assess awareness separately for each persona, but rather only assessed awareness as generally associated with an "Asian American accent" used by *some* Asian Americans. In this case, although the experimental task aimed to assess behavioral outcomes based variable associations with particular personae, we cannot separate the association of a particular variable with a particular personae in the gradient assessment of awareness. Thus, the relevance of an individual's awareness score to their behavior in a particular persona pairing condition may have been variable. These are both factors that may have reduced the effectiveness of the composite awareness score as a predictor of behavior.

Another limitation, linked to the first, is the sample size for individual-level analysis. If effect sizes are rather small in reality, a larger sample size would help to detect an effect of awareness with more certainty. The current sample size is on the lower end of typical individual differences studies, but provides a start that indicates potential patterns in the predicted direction. More data can be collected to reduce this uncertainty.

To further explore these questions, there are many avenues for future research to take. For example, a similar methodology could be used to measure awareness for multiple variables and individuals, as this study did, in order to directly compare how awareness may be related to strength of perceptual behavior. Depending on the particular sociolinguistic contexts under study, other types of conditions that could be included for comparison are communities that may be expected to have different levels of awareness, as well as types of linguistic items that are differently tied to conscious access (e.g., stereotyped phrases and words). In the AAC context, for example, future research could additionally investigate awareness and sociophonetic associations of other aspects of speech (e.g., longer VOT in stops, f0) in locations varying in population concentration of AACs (i.e., where lower exposure may be linked to lower awareness at the community level). Alternatively, using an approach to manipulate awareness within an experimental context would be an even more targeted way to test the role of awareness in sociophonetic perception.

6.6.4 Conclusion

Taken together, this study provides evidence that /oʊ/-backing and /ð/-stopping are sociophonetic variables linked to certain AAC personae, offering a number of new insights into Asian American speech and ideologies about "sounding Asian". With regards to questions about sociophonetic cognition, in addition to contributing evidence on the role of personae as a perceptual construct, it also finds support for awareness as a mediating factor in the encoding of sociophonetic representations.

CHAPTER VII

General Discussion & Conclusions

7.1 Summary of Results

As presented in Chapter I and motivated in Chapter II , the goals of this dissertation were: (a) to investigate the influence of ideology on sociophonetic representations, perception, and processing; and (b) to contribute empirical evidence from production, perception, and metalinguistic commentary to our understanding of AAC speech. The three guiding questions were:

1. How is AAC speech ideologically constructed, cognitively represented, and perceptually processed?
2. How do social personae factor into sociophonetic representations and perceptual processing?
3. Does sociolinguistic awareness mediate the strength of sociophonetic representations and perceptual processing?

I attempted to address these questions through a series of studies sampling Asian Americans from California—a sample that was expected to provide a range of experiences with AAC speech and related ideologies.

The exploratory production study (Chapter III), conducted via a YouTube-based corpus analysis, revealed much individual-level phonetic variation in

Californian Asian American speakers with regards to /oʊ/-backing, /oʊ/-monophthongization, and speech rhythm. Of these features, /oʊ/-backing was the most consistently found in this sample of AAC speech compared to benchmark non-AAC speech. It was especially notable for Chinese and Korean American speakers on the whole, as well as for two individual speakers, one Chinese American and one Korean American, who clustered together based on similar patterns of extreme /oʊ/-backing. This result implies that /oʊ/-backing could be a feature that exists in both ideological constructions and cognitive representations of AAC speech, based partially on direct or interactional exposure.

The exploratory perception and ideology study (Chapter IV) combined an auditory social evaluation task with an ideology-focused questionnaire to address this possibility. Two additional aims of this study were to (a) identify other potential perceptual cues to AAC speech along with their relative degree of awareness, and (b) describe ideological social types of AACs prevalent in this context. Based on naturalistic speech samples from the YouTube corpus, Californian Asian American listeners were able to consistently identify AAC speakers on the whole as Asian (though they were not accurate at identifying specific ethnicity), and speakers with a greater degree of /oʊ/-backing were rated as more likely to be Asian. /oʊ/-backing co-occurred with /ð/-stopping in the speech perceived as most likely Asian, and these were also the two features rated as most familiar and accurate in terms of the ideological link to an "Asian American accent". Open-ended metalinguistic commentary offered explicit discussion of the ideological and perceptual connection between hearing /oʊ/-backing or /ð/-stopping and AAC identity. These results suggest that /oʊ/-backing and /ð/-stopping are relevant features in ideologies of AAC speech as well as specifically leveraged by listeners in perceptual ethnic identification of AACs. Relatively speaking, /ð/-stopping was rated higher on both familiarity and accuracy compared to /oʊ/-backing, which I interpreted as indicating slightly more community-wide awareness for /ð/-stopping as a feature linked to AACs.

Ideological commentary about types of AACs further confirmed that Californian Asian Americans recognize various subcategories of AACs with distinct styles, including linguistic styles. The most prominently described types included a "FOB" Asian American (characterized in part by foreign-accentedness) and a "whitewashed" Asian American. Following Pyke and Dang (2003), I interpret these ideological constructs of AAC types as existing on a continuum of acculturation—from an ethnic orientation to a mainstream orientation—in the minds of co-ethnics. Categories on this continuum are further cognitively linked to broad constructs of foreign Asian accents, Asian American accents, and white American accents.

In a series of social evaluation surveys intended to more concretely capture ideologies of differently-acculturated AACs (Chapter V), I identified photos of people that represent emergent social types involving a relatively ethnically-oriented Asian American (interpreted as a FOB Asian persona) and a relatively mainstream-oriented bicultural Asian American (interpreted as a Normal Asian persona). Based on visual style information alone (e.g., clothing, hair, makeup), ethnically-oriented faces were linked to expectations of speech that was relatively foreign-accented, non-native, and unclear while mainstream-oriented faces led to expectations of relatively American-accented, native, and clear speech.

Voice-based social evaluations converged with findings from the previous study that listeners consistently attribute Asian racioethnic identity to certain AAC speakers, even in relatively controlled stimuli. Beyond that, it clearly showed that particular speakers were identified based on speech alone as specifically Asian Americans who grew up in North America. My interpretation is that these AAC speakers' speech style conform to ideological constructions of an "Asian

American accent" while other AAC speech does not. Trends in face-to-voice mapping results hinted that voices may also be able to sound specifically like a mainstream-oriented AAC compared to an ethnically-oriented AAC. Together, these results complement growing evidence that the persona is a relevant layer of social categorization in cognitive representations that mediates linguistic expectations, in this case of broad accent constructs at a microsocial level.

Since /oʊ/-backing and /ð/-stopping were excluded from these speech samples, this indicates that they are not necessary to inducing an Asian or local AAC social impression. However, they could still be part of the full set of phonetic features ideologically and cognitively linked to AAC identity, or to specific AAC personae differing on perceived cultural orientation, that contribute to such an impression.

The following study (Chapter VI) directly tested the role of personae and awareness in sociophonetic perception while also assessing the sociophonetic indices of /oʊ/-backing and /ð/-stopping. Employing a two-alternative forced choice social categorization task with an ideology questionnaire, I found that /ð/-stopping was self-reported with a higher degree of awareness than /oʊ/-backing was, and at the group level led to a larger effects of behavioral selection bias for Ethnic Asian persona photos (vs. Mainstream Asian or Mainstream White). At the individual level, trends of greater awareness being related to larger effect sizes were in the predicted direction but small and uncertain. Overall, I interpret this as support for the mediating role of ideological awareness and personae in cognitive representations such that greater awareness increased the relative strength of weights on the indices between /ð/-stopping and social information about ethnically-oriented AACs.

In the same task, /oʊ/-backing led to smaller effects of selection bias for the Mainstream Asian (vs. Mainstream White) and Ethnic Asian (vs. Mainstream

Asian or Mainstream White) photos. In other words, /ð/-stopping led to greater selection of Ethnic Asian photos but not Mainstream Asian photos while /oʊ/-backing led to greater selection of both. This suggests that the former may be indexed only to a particular AAC persona instead of a broad macrosocial Asian or AAC category while the latter may be indexed to both of the investigated personae, Asian identity, and/or AAC identity. Either way, this result highlights the importance of capturing microsocial and/or social persona information to provide a fuller and potentially more cognitively realistic view of sociophonetic representations and socioperceptual processing.

7.2 Ideology in Sociophonetic Cognition

7.2.1 Nature of Sociophonetic Representations

The results of this dissertation contribute to discussion of what the social categories or social information linked to phonetic detail *are*, as well as what kinds of categories or information are used in tasks relating to socioperceptual behavior. Many previous approaches to socially-informed expectations and perceptual processing have focused on categories at the macrosocial level, discussing perceptual behavior as mediated by social categories like race/ethnicity, age, gender, national identity and more (e.g., Strand, 1999; Koops et al., 2008; Staum Casasanto, 2010; McGowan, 2015). But even a "simple" macrosocial category is embedded in ideologies. This has been discussed as stereotypes of a particular identity that influence perceptual expectations (e.g., Niedzielski, 1999). In the current sociolinguistic context, it is clear that "Asian" seems to be stereotypically associated with foreignness (Chapters IV and V), also recognized by studies exploring the effect of this category on perceptual processing and speech comprehension (Rubin, 1992; M. Babel & Russell, 2015; McGowan, 2015). In light of this, I would suggest that the characterization of behavior driven by expectations of a macrosocial category of people could potentially be reframed—especially at an individual cognitive level—as driven by the *idea* of that macrosocial category, which could be a particular persona or stereotypical character. In an exemplar theoretic approach, we can recast discussion of exposure-based exemplars into one about ideology-guided exemplars, in which experience includes exposure but also other forms of sociolinguistic knowledge (Drager & Kirtley, 2016).

In terms of personae, the results of Chapters V and VI align with research that supports the role of persona-based information in influencing sociophonetic perceptual behavior, including social evaluation or categorization tasks as well as linguistic perception tasks (e.g., MacFarlane & Stuart-Smith, 2012; D'Onofrio, 2018a, 2018b, 2019). These can include locally relevant imagined identities, like "Glasgow Uni" and "Glaswegian" (MacFarlane & Stuart-Smith, 2012), and also more generalized ideas like "Valley Girl" and "Business Professional" (D'Onofrio, 2018b). These personae, as well as "FOB Asian" or "Normal Asian", could be considered microsocial categories. This indicates that microsocial information is related to speech and cognitively accessible to listeners during perceptual tasks. However, more research will be needed to fully understand how microsocial information is utilized by individual listeners in sociolinguistic perceptual processing (see also D'Onofrio, 2018a). For example, the FOB Asian and Normal Asian personae may be conceptually subsumed by broader "Asian American/Canadian" and "Asian" ideological constructs. One relevant question might be whether narrower subordinate categories are necessarily the basis of social perceptions where available, or framed more generally, how might this type of social information interact with overlapping categories of various levels of abstraction in representation and processing?

In order to examine the role of personae, I employed an auditory social categorization task with multiple speakers and multiple photos. Previous studies have used images of brand logos (MacFarlane & Stuart-Smith, 2012), illustrations (Leigh, 2021), and single photos (D'Onofrio, 2019) to represent personae. Finding similar persona-based effects through the use of multiple naturalistic photos of faces and a variety of voices adds an element of generalizability and ecological validity to the sociophonetic perception literature on this topic. That is, the current results may help to further link this set of findings to realistic social interaction settings where speakers' social identities and linguistic expectations are inferred via visual details.

The outcomes of this dissertation with regards to personae also highlight some methodological considerations for future sociophonetic perception research. Minimally, it will be important to assess and acknowledge the stylistic details of the stimuli, as even supposedly controlled stimuli are portraying certain impressions to participants. In other words, we want to better understand what perceivers are actually interpreting as the social information when given a social category or image intended to represent that category. Reporting contextual details would help provide a clearer picture of the extent that a finding may be generalizable.

7.2.2 Strength of Sociophonetic Associations

This dissertation's findings are additionally consistent with proposed roles of awareness, salience, and attention in an exemplar-based model of sociolinguistic cognition. Conceptualized as a feedback loop (e.g., Günther et al., 2017), awareness of sociolinguistic variation (e.g., /oʊ/-backing and /ð/-stopping as produced by certain AACs) based on past experiences should tend to increase strength of expectations when encountering relevant speech. If expectations are met (i.e., hearing backed /oʊ/ or stopped /ð/ from an AAC), this can result in salience, therefore drawing attention to those sociophonetic variants and leading exemplars and social indices of those experiences to be stored with stronger weights in long-term memory. Encountering sociolinguistic variation metalinguistically, in discussions or thoughts about such associations (e.g., some AACs say “th” like “d”), may also lead to storage or strengthening of exemplars and indices via activation (e.g., Drager & Kirtley, 2016). Exemplars with stronger weights accumulated over time will then activate more quickly and readily, allowing for the possibility of stronger activation of associated information than another set of exemplars with an equal or greater number of exemplars but weaker prominence (e.g., Sumner et al., 2014).

If so, and if all else were equal, we would expect the presence of variables with a higher degree of average awareness to result in larger socioperceptual behavior effects, whether making social or linguistic decisions. That was the result found in this dissertation, having estimated awareness gradiently via questionnaires about experiences with these sociophonetic associations: Compared to /oʊ/-backing, /ð/-stopping in AAC speech was self-reported as more frequently encountered in metalinguistic and linguistic experiences as well as more subjectively familiar, and led to larger selection biases for ethnically-oriented AAC photos (Chapter VI). Drager and Kirtley (2016) also predicted that this relationship should hold at the individual level, which there were trends for, but not enough data to determine with certainty.

The current approach adds to the mixed literature on this topic as well as contributes methodology for conceptualizing and measuring awareness. Past studies that did not directly assess awareness or salience did not necessarily find relevant effects (e.g., Lawrence, 2015). On the other hand, this dissertation joins Juskan (2018)—in which social salience for variables was assessed holistically based on multiple sources of information, including prevalence in open-ended metalinguistic commentary—in finding a predicted behavioral effect dependent on degree of awareness or salience. Future work examining and comparing a wider range of variables at once, like Juskan (2018) did, in other contexts would allow for confirmation of results such as these. For example, in the AAC context, examining perceptual behavior in response to variables with either no associated awareness or relatively low awareness (e.g., vocal fry, according to the results of Chapter IV) compared to variables with relatively high awareness (e.g., /ð/-stopping) could be a fruitful endeavor to further test the influence of awareness.

One intersecting aspect that was not fully addressed in the current dissertation is frequency of exposure. While not the sole factor in determining sociophonetic association strength, exposure is the basis of exemplar-based models which ideological factors like awareness may act on. If exposure frequency were equal between two sociolinguistic variables, we would expect degree of awareness to straightforwardly mediate differences in association strength. However, if exposure was unequal, the result could either be exaggerated differences in socioindexical association strength (in the case of high exposure along with high awareness), or minimized differences (in the case of low exposure paired with high awareness). Thus, degree of difference in awareness may not translate directly to degree of difference in perceptual behavior. In the extreme, there could even be a possibility that much higher exposure with lower awareness may result in stronger overall association strength than much lower exposure with higher awareness. While there is some evidence of /oʊ/-backing in AAC speech production (Chapter III), /ð/-stopping has not been specifically studied in the same context, other than some hints from post-hoc analyses (e.g., A. Cheng & Cho, 2021) and metalinguistic commentary. Even if it were, the prevalence of any particular variant in individual experiences is not necessarily easily measured. Still, to build a more complete picture of the role of awareness in the AAC context and any future studies, it will be important to better understand patterns of exposure along with awareness. One idea might be to examine variables with different combinations of exposure and awareness to assess the contribution of each. Another approach, which was taken in this dissertation, would be to focus on individual awareness for the same variable, assuming roughly similar exposure across a population.

Unlike previous studies (e.g., Juskan, 2018), I focused on quantifying awareness directly and gradiently at an individual level in an attempt to capture more nuanced variation in awareness and salience beyond prevalence of open-ended comments at a group level. When measured this way, I found gradient degree of awareness across individuals that spanned the full possible range for both /oʊ/-backing and /ð/-stopping. This gradience was uncovered despite the fact that most people did not comment on /oʊ/ as a notable feature of the speech samples they heard, even for those containing extremely backed /oʊ/ (Chapters IV), and no respondent specifically highlighted /oʊ/ when asked to describe features of AAC speech, even if they referred to pronunciation of vowels at a vague level (Chapters IV and VI).

Taken together, the gradient measures and metalinguistic responses suggest that the presence or prevalence of explicit discussion of a particular sociophonetic association may not be a particularly accurate or precise indicator of awareness. That is, while it can be taken as positive evidence of some form of awareness, it does not capture much information and may be somewhat misleading. On the one hand, awareness of a variable can exist even if respondents do not bring it up in explicit commentary, possibly because it was relatively low in awareness and/or it did not happen to be accessible at that particular moment. On the other, even if respondents do tend to explicitly mention a variable, this association can be tied to varying degrees of awareness. Altogether, this dissertation calls attention to the value of assessing instead of inferring degree of awareness (or social salience). Doing so could potentially help to explain some of the null findings in the sociophonetic perception literature, if degree awareness or salience was in fact relatively low at an individual level, or at least rule out that explanation to help identify the actual reason.

However, further refinement of measures and tests of awareness, especially at the individual level, is needed. In Chapter IV, I employed two subjective Likert scale questions—one for familiarity and one for accuracy—that were interpreted together as targeting likely degree of conscious access to an ideological sociophonetic association. In Chapter VI, I added additional Likert scale questions about amount of subjective exposure, but the extent to which each question may be important to the construct of awareness was not specifically validated. As such, although I constructed a simple unweighted linear combination of subjective familiarity and exposure responses as a first pass to estimate degree of awareness, this method may not be sufficient to capture the construct in detail. Moreover, the measures I used were not specific enough. Awareness estimates for particular sociophonetic variants were assessed only for their association with AACs (framed as an AAC accent) rather than each AAC persona, even though I tested behavior at the persona level. It may be unsurprising, then, that only a small potential behavioral trend was found for individuals, especially given the likely insufficient sample size. Individual variation of other interacting sociocognitive traits, such as tendency to learn or apply sociolinguistic associations, may also reduce the ability to detect an effect of awareness without a larger sample (e.g., G. J. Docherty et al., 2013; Rácz et al., 2020; L. S. P. Cheng, 2021). Regardless, the steps taken here are promising, and future work can build on the current methods to create measures of awareness that more accurately and precisely capture the construct in order to better assess its true impact on sociophonetic perception.

Although awareness and salience may lead to reinforced associations over time, the current dissertation does not directly probe how sociophonetic associations are originally formed or learned, nor how they become strengthened variably over time. I touch on two points related to these aspects of the current framework. One question of potential interest is how different types of salience may play into why certain phonetic variables come to be socially indexed or consciously accessible in the first place. Is /ð/-stopping more salient in the perceptual domain based on stimulus-inherent phonetic properties compared to other features, therefore allowing it to more easily become socially indexed to AACs in awareness (e.g., see Rácz, 2013; Günther et al., 2017)? Could /ð/-stopping be more easily attached to AAC identities with greater awareness because it is also tied to other ethnic groups in broader sociolinguistic experiences, leading to greater shared social salience (e.g., G. J. Docherty et al., 2013)? Another relevant question is whether increased awareness can be causally linked to an increase in sociophonetic association strength, more directly supporting the proposed relationship between awareness and sociophonetic association strength. Generally, these questions could be experimentally tested via intervention or learning studies, such as designs that manipulate the presence or absence of metalinguistic comments in between perceptual tasks, or designs that compare learning of artificial novel social categories for linguistic variants deemed to be salient in different ways. Overall, it will likely be useful to pursue future investigations of awareness and salience in different sociolinguistic variables and contexts, including artificial ones, as different combinations of factors can be examined or manipulated, providing new insight to further develop our models of sociophonetic representations and perception.

7.3 Ideology and AAC Speech

7.3.1 Perceptual Cues to AAC Speech

This dissertation moves us closer to the goal of understanding AAC speech—specifically from the viewpoint of what is *perceived* as AAC speech, sourced from linguistic ideologies. At a high level, the outcomes of AAC ethnic identification and social evaluation (Chapters IV and V) are consistent with general expectations from the literature. First, the results closely mirror findings from past research on perceptual ethnic identification of AAC speech (Hanna, 1997; Newman & Wu, 2011; P. Wong & Babel, 2017; Nagy et al., 2020; A. Cheng & Cho, 2021): Asian American listeners were able to identify some but not all AAC speakers as Asian—and specifically Asian from North America—but they were not necessarily able to identify specific ethnic identity. Second, the results provide quantitative evidence that Asian identity, as inferred through speech, may be generally associated with foreignness (Tuan, 1998; Reyes & Lo, 2009), and beyond that, a FOB Asian persona is specifically associated with foreign-accentedness while a Normal Asian persona is not (Pyke & Dang, 2003).

More specifically, the results offer novel, experimental evidence that /oʊ/-backing and /ð/-stopping are relevant sociophonetic variables linked to some forms of AAC identity. They appear to be present in linguistic ideologies and cognitive representations, influencing perceptual behavior in social evaluation and categorization tasks. /ð/-stopping may be tied with more awareness to "Asian" and "FOB Asian" social categories, linked to foreignness, while /oʊ/-backing may be additionally tied to a "Normal Asian" category that is not itself linked to foreignness, with overall lower amounts of awareness. There is opportunity for a deeper dive into the indexical fields of these variables, including how different socioindexical associations may interact with impressions of AAC identity.

Future research can continue this work by systematically examining the role of other linguistic features to fill out the picture of what constitutes perceptual cues to AAC speech. A focus on suprasegmental phonetic features will likely be important. Based on ratings of familiarity and accuracy (Chapter IV) other features possibly represented in the ideological construct of AAC speech are higher pitch and quieter speech. Across studies, a large proportion of open-ended metalinguistic comments about an "Asian American accent" also use terms like "tone", "intonation", "inflection", and "pitch", or mention other relatively holistic elements of a voice. One respondent suggested that this phenomenon is "not an accent and more of the general tone/pitch of our voices", implying a prominent role for voice-related characteristics for some listeners; related comments include ones differentiating AAC speech with identifiable pronunciation differences (e.g., "lisp", "accent") and AAC speech with "standard pronunciation" but which still gives off the impression of an AAC speaker. Potentially related subjective descriptions of AAC speech include sounding "calming and smooth" rather than "energetic", as well as impressions of a "gentleness" or "softness" to AAC voices—identifying what these impressions refer to in the speech signal could help to better understand the perceived "Asian American voice" and how it may differ from an "Asian American accent."

Additionally, ideological commentary points towards the potential for lexical, morphological, or syntactic cues to AAC identity (e.g., deletion of plural morphemes, constructions like "more cleaner" rather than "cleaner"). In one case, a respondent indicated that they sometimes "can tell if someone is Asian American or more culturally Asian" by the way they write on online posts, even without any "spelling/grammatical errors", while another suggested that they noticed structural contact where "vocabulary/sentence structure" may appear to be a "[more direct] translation of the grammar of Asian languages." As far as I know, these types of variation have not been studied with regards to AAC speech but could be a relevant part of the puzzle.

In this dissertation, I only examined single features at a time, not what happens when multiple variables occur together. In Chapter IV, various features co-occurred in the naturalistic speech samples, including /oʊ/-backing and /ð/-stopping in the speakers most commonly perceived as Asian, but I only systematically assessed the presence of one. Since the presence of both features were found on average to mediate increased responses to an Ethnic Asian persona (Chapter VI), it could be reasonable to assume that both features contributed additively to create a strong (ethnically-oriented) AAC impression (e.g., via activation of those social categories), where /ð/-stopping may have contributed more due to its greater awareness. Future research could directly test this question.

Many other questions remain in terms of how exactly these and other features in the signal may combine to form a particular Asian or AAC impression. One question is how the strength of each cue (e.g., extremely backed /oʊ, weakly stopped /ð/) may be linked to particular AAC social personae. A second question is how other features, like suprasegmental or voice cues, may interact with cues like /oʊ/-backing and /ð/-stopping, including whether some cues are necessary to an AAC impression while others only add certainty to that impression in combination with other existing cues. For example, voices presented without instances of /oʊ/-backing and /ð/-stopping were rated variably on likelihood of being Asian (Chapter V), and photo selection bias effects also differed in strength between voices (Chapter VI). Though there were other differences between voices in that task as well (e.g., how they produced the target variants), it is plausible that part of the by-speaker variation could be attributed to interactions between the target features and other existing features of the voice.

Where do these ideologies and representations come from? One possible source of /oʊ/-backing associations is production, as seen in Chapter III. The finding that /oʊ/-backing was produced by ethnic Korean speakers aligns with prior production studies (e.g., L. Jeon, 2017; A. Cheng, 2020), though the result for ethnic Chinese speakers runs somewhat contrary to reports of relative /oʊ/-fronting for this group in California (Hall-Lew, 2009; A. Cheng et al., 2016). Nevertheless, the Chinese group in A. Cheng et al. (2016) did show only a moderate pattern of /oʊ/-fronting, therefore produce a slightly more backed /oʊ/ than the white group (though a less backed /oʊ/ than the Korean group); this contrast may have contributed to a pan-ethnic or multi-ethnic association with AACs in this context. /ð/-stopping was not examined in the current dissertation, but it was found in production of speakers most identified as Asian in previous perceptual studies of Korean American and Chinese Canadian speakers (P. Wong & Babel, 2017; A. Cheng & Cho, 2021), potentially suggesting that it occurs in ethnically-oriented AAC speech. This will need to be studied further to confirm the prevalence of /ð/-stopping in AAC production. Respondents also suggested an ideological link between this variable and Vietnamese American speaker production in particular, which could be of potential interest for future research as well.

Beyond direct exposure to AAC speech, experience with contact varieties of Asian English speech may add to the cluster of features available in the ideological space surrounding AAC as a category, linked to Asian identity, a FOB Asian persona, or related social categories. Similarly, both features may be associated with stereotyped portrayals of foreign Asian-accented speech in comedic performance (Chun, 2004), which may play a role in creating or bolstering those associations, potentially through increasing awareness. In addition, the concepts of /oʊ/-backing and /ð/-stopping may arise in metalinguistic commentary. In Chapter IV participants reported that they were familiar with the statements describing these features as part of an Asian American accent, even if they did not agree, and many also indicated that some amount of people talk about the Asian American accent. In Chapter VI, many participants indicated that they came across the accent, /oʊ/-backing, or /ð/-stopping in their daily life to some extent, including possible metalinguistic discussion. This type of experience, as well as possible imitation of AAC speech, may have contributed to either forming or reinforcing associations between these features and the social category of AAC. Future work may seek to investigate ideological sources of these associations and how features linked to the various related social categories in this context may relate to each other.

7.3.2 Variation in AAC Speech

This dissertation demonstrates a particular methodological approach to a relatively unknown context of sociolinguistic variation, which, in this case, was sourced from the observation that listeners could identify AAC speakers as such. Using the ideological conception of "sounding Asian" as the focal point, I first began with exploratory studies involving naturalistic speech data and a broader scope of inquiry, conducting both qualitative and quantitative analysis to build up an understanding of production, perception, and ideologies in this context. Then, I moved on to more targeted studies using controlled speech stimuli focused on particular variables, describing sociolinguistic associations in visual and auditory domains before putting it all together in an experimental task. However, associations and ideologies relating to AAC speech can be expected to vary in many more ways than examined here, including based on factors such as region and specific ethnicity. The scope of this research across all studies was limited to Asian Americans from California, and the main self-reported respondent ethnic groups tended to be Chinese, Vietnamese, and Filipino. I consider below a few of the implications.

On the production side (Chapter III), features used to index AAC identities are embedded in the local context and, even if some features are shared across regions, this may lead to differences in awareness and salience. For example, /oʊ/-fronting is a known feature of Californian speech, seeming indexed to California, Valley Girl, and whiteness (Chapter IV). It may be that against this backdrop, /oʊ/-backing becomes especially salient in opposition, resulting in increased prominence as an AAC sociophonetic variable in production and/or perception. That is, if expectations (based on long-term memory) of hearing /oʊ/-fronting in California are strong, then experiencing a violation of those expectations when /oʊ/ is backed (or less fronted) would be predicted to draw much attention to /oʊ/-backing and result in exemplars stored with strong weighting. In another location where /oʊ/-fronting is less present or socially marked, and therefore not particularly expected in speech, /oʊ/-backing may neither be salient nor leveraged as an ideological feature of AAC speech, even if it happened to occur in experiences to the same extent.

Existing ethnic identification studies have tended to focus on matched speaker and listener populations, though located in different regions. That a similar pattern of results were found for local populations' ability to identify local AACs by speech in different places—such as New York (Newman & Wu, 2011), Vancouver (P. Wong & Babel, 2017), and California (Chapter IV)—is interesting, but it is possible that listeners in each location relied on different perceptual cues relevant to their context alone. In Chapters V and VI, while listeners were always Californian, the speakers were of a few ethnic groups from various locations across North America (with some spending time in Asia first). Some were able to elicit similar impressions of AAC identity, especially one who grew up in North America, which more directly indicates that listeners were able to access generalized ideologies of AAC speech; these ideologies potentially include a multi-ethnic collection of features that at least partially transcends experience with local varieties of specific ethnic groups (see also A. Cheng & Cho, 2021).

At the same time, people have variable experiences and level of familiarity with AACs from particular regions and ethnic groups; these together should contribute to variation in their cognitive representations of AAC speech. As we might expect, previous results show that individuals with greater familiarity seem more likely able to identify a speaker as a specific ethnicity (e.g., P. Wong & Babel, 2017; A. Cheng & Cho, 2021). There are likely to be systematic differences in the most relevant perceptual cues to AAC identity for individuals with different patterns of ethnic experience, not to mention individual differences in which features become most salient to them. The same patterns of experience may also be expected to determine the social categories in their cognitive representations, such as whether they have models of Korean American speech separate from Chinese American and Vietnamese American speech, if they group these together under AAC speech, or both. This could also be applied to whether listeners have categories differentiating East Asian speech from Southeast Asian speech, as well as any other socially relevant distinction. In general, understanding the exact representations that an individual brings to a sociophonetic perception task involving AAC speech, and which is implicated in their expectations and processing, may be complex and will require further refinement.

Along the same lines, there is a need to better understand other types of intersectional AAC categories as linked to ideologies and associations, and how this influences sociolinguistic perception. For instance, across all the studies in this dissertation, only people who were (perceived as) AAC women were included in the auditory and visual stimuli. This was done partially due to constraints of speaker selection and speaker recording, but also served to reduce an axis of variation expected to influence perceptions of AAC speech. Differences have been suggested in previous work with regards to how differently-gendered AAC voices may be identified (A. Cheng & Cho, 2021) and evaluated (Bauman, 2013). The existence of ideological gender differences within AAC communities was further hinted at by metalinguistic commentary in this dissertation. For example, descriptions for AAC women's voices were explicitly differentiated from AAC men's voices by some participants. Descriptors such as "sweetness" or "femininity" were attributed to the speech of AAC women, and there were also many references to Valley Girl speech styles. More broadly, some respondents explicitly identified the existence of multiple Asian American accents which may be tied to different locally-relevant types of people, further indicating that there is not a single way to "sound Asian American." Whether the current findings, such as the roles of /oʊ/-backing and /ð/-stopping, will necessarily apply to speakers from other subgroups of AACs remains to be seen.

Finally, given that the participant pool included only Asian Americans, a major question for the future will be how these results may speak to ideologies, representations, and perception in other populations. D'Onofrio (2019), for example, shows that a largely non-Asian American population appear to both distinguish and hold different linguistic expectations for a less American-perceived Asian photo compared to a more American-periceved Asian photo at the broad level of foreign- or mainstream-accented speech. However, the details of this distinction are unclear, including whether /oʊ/-backing and /ð/-stopping may be used by different populations in the same way or to the same extent. Put another way, might different populations' exposure to and awareness of these particular features and social categories vary systematically, resulting in a different configuration of cognitive representations (e.g., due to group-level differences in interactional exposure, mock performance, and metalinguistic commentary of Asian foreign-accented or AAC speech)? For example, in this dissertation, /ð/-stopping appeared to be tied to foreign-accented speech and a FOB Asian persona while not being associated with a Normal Asian persona (Chapter VI), but it is possible that a different population may associate /ð/-stopping with a broader set of Asian and AAC identities such that the presence of stopped /ð/ could elicit expectations of someone that might be represented by a Normal Asian persona as well. Another possibility could be that a similar pattern of results is found with different magnitudes, such as a weaker effect for /oʊ/-backing or a stronger effect for /ð/-stopping, which would suggest varying strengths of these particular associations given varied experiences and ideologies. Studies targeting different populations will be needed to explore the generalizability of these results beyond Asian Americans from California.

In general, the current dissertation contributes insights to a goal of understanding how AACs may be sociolinguistically perceived in realistic social contexts. Much of the research involving the influence of seeing Asian faces on the interpretation and processing of speech aim to link these findings to potential real-world consequences, such as in job seeking or university teaching situations (e.g., Rubin, 1992; Y. Zheng & Samuel, 2017). In light of persona-based findings, a question we can ask is how well research investigating the interaction of speech and macrosocial racialized categories in fact connect to real-world situations such as those, and for which Asian-racialized individuals. Individuals who express themselves and their identity in different ways—both visually and auditorily—can potentially elicit very different attitudes, expectations, and potential speech processing consequences, even if they are racialized the same. Future research may consider exploring these differences, including in applied contexts. A basic question might ask whether hearing AAC speech that can be identified as an Asian North American individual results in particular perceptual impressions or behavior that differ from hearing Asian foreign-accented speech or speech perceived as non-AAC. Delving deeper into the interactions of AAC-specific stylistic parameters—such as looking or sounding foreign, ethnically-oriented, or mainstream-oriented—with other aspects of style would help better characterize the diverse social and linguistic experiences of different AAC individuals, including nuanced consideration of experiences of linguistic discrimination.

7.4 Conclusion

This dissertation highlights the importance of accounting for ideology both in the context of AAC speech and in models of sociolinguistic cognition more generally. Presence of /ð/-stopping and /oʊ/-backing in speech were both found to increase impressions of AAC identity, but did so for different social types and to different degrees. The complex ideological landscape of AAC speech will need to be acknowledged and addressed to build a more complete picture of AAC sociolinguistic experiences. Personae and awareness were found to mediate sociophonetic perceptual behavior, suggesting that the types of social categories in cognitive representations and the strength of representational links between social and linguistic information are guided by ideologies. This aligns with exemplar theory approaches that foreground salience and attention in the encoding and reinforcement of linguistic experiences in memory. Adopting a framework where ideology and individual interpretation of experiences underlie sociolinguistic representations and perception moves us closer to an understanding how speech is processed in the social world.

APPENDICES

APPENDIX A

Social Evaluation Surveys

Survey 1b: Instrument

Please describe your impressions of this person using three words. You can use whatever adjectives you like, but try not to focus just on physical attributes, or what the person is wearing.

How old do you think this person looks? (Multiple selections are possible.)

- ❏ Teens
- ❏ 20s
- ❏ 30s
- ❏ Over 40

What do you think this person's occupation is?

What activities or hobbies do you think this person likes to do?

Where do you think this person grew up?

If you had to guess, what do you think this person's ethnic background is?

If you had to guess, how do you think this person would speak?

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

Casual	❍	❍	❍	❍	❍	❍	❍	Formal
Feminine	❍	❍	❍	❍	❍	❍	❍	Not feminine
Cool	❍	❍	❍	❍	❍	❍	❍	Uncool
Nerdy	❍	❍	❍	❍	❍	❍	❍	Not nerdy
Clear speaker	❍	❍	❍	❍	❍	❍	❍	Unclear speaker
Slow speaker	❍	❍	❍	❍	❍	❍	❍	Fast speaker

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

Unattractive	❍	❍	❍	❍	❍	❍	❍	Attractive
Unintelligent	❍	❍	❍	❍	❍	❍	❍	Intelligent
Not confident	❍	❍	❍	❍	❍	❍	❍	Confident
Not enthusiastic	❍	❍	❍	❍	❍	❍	❍	Enthusiastic
Unlikeable	❍	❍	❍	❍	❍	❍	❍	Likeable
Unfriendly	❍	❍	❍	❍	❍	❍	❍	Friendly

American	❍	❍	❍	❍	❍	❍	❍	Not American
Aligned with American culture	❍	❍	❍	❍	❍	❍	❍	Aligned with ethnic culture
Native speaker of English	❍	❍	❍	❍	❍	❍	❍	Non-native speaker of English
American-accented	❍	❍	❍	❍	❍	❍	❍	Foreign-accented

Do you recognize this person? If so, who do you think they are?

❍ No

❍ Yes [ ]

(optional) Did anything else about this person stand out to you?

[ ]

Survey 2: Instrument

Please describe your impressions of this person based only on their voice, using up to three descriptors. You can use whatever adjectives you like, but try not to focus just on physical attributes. It's OK if you're not sure, just type what comes to mind.

How old do you think this person sounds? (Multiple selections are possible.)

- ❏ Teens
- ❏ 20s
- ❏ 30s
- ❏ Over 40

Where do you think this person grew up?

If you had to guess, what do you think this person's ethnic background is?

How would you describe the way this person speaks and/or how their voice sounds?

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

Casual	❍	❍	❍	❍	❍	❍	❍	Formal
Feminine	❍	❍	❍	❍	❍	❍	❍	Not feminine
Cool	❍	❍	❍	❍	❍	❍	❍	Uncool
Nerdy	❍	❍	❍	❍	❍	❍	❍	Not nerdy
Clear speaker	❍	❍	❍	❍	❍	❍	❍	Unclear speaker
Slow speaker	❍	❍	❍	❍	❍	❍	❍	Fast speaker

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

Unattractive	❍	❍	❍	❍	❍	❍	❍	Attractive
Unintelligent	❍	❍	❍	❍	❍	❍	❍	Intelligent
Not confident	❍	❍	❍	❍	❍	❍	❍	Confident
Not enthusiastic	❍	❍	❍	❍	❍	❍	❍	Enthusiastic
Unlikeable	❍	❍	❍	❍	❍	❍	❍	Likeable
Unfriendly	❍	❍	❍	❍	❍	❍	❍	Friendly

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

American	❍	❍	❍	❍	❍	❍	❍	Not American
Aligned with American culture	❍	❍	❍	❍	❍	❍	❍	Aligned with ethnic culture
Native speaker of English	❍	❍	❍	❍	❍	❍	❍	Non-native speaker of English
American-accented	❍	❍	❍	❍	❍	❍	❍	Foreign-accented

Labels for the following questions:

- Extremely unlikely (1)
- Unlikely (2)
- Somewhat unlikely (3)
- Neutral (4)
- Somewhat likely (5)
- Likely (6)
- Extremely likely (7)

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

	1	2	3	4	5	6	7
Grew up in the US/Canada	❍	❍	❍	❍	❍	❍	❍
Grew up in Europe	❍	❍	❍	❍	❍	❍	❍
Grew up in Asia	❍	❍	❍	❍	❍	❍	❍

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

	1	2	3	4	5	6	7
Grew up in California	❍	❍	❍	❍	❍	❍	❍
Grew up in the Midwest	❍	❍	❍	❍	❍	❍	❍
Grew up in the Southern US	❍	❍	❍	❍	❍	❍	❍
Grew up elsewhere in the US or in Canada	❍	❍	❍	❍	❍	❍	❍

Please rate this person on the following characteristics. It's OK if you're not sure, just give your best guess.

	1	2	3	4	5	6	7
Asian or Asian American	❍	❍	❍	❍	❍	❍	❍
Black or Black American	❍	❍	❍	❍	❍	❍	❍
Latin/Hispanic or Latin/Hispanic American	❍	❍	❍	❍	❍	❍	❍
White or White American	❍	❍	❍	❍	❍	❍	❍

(optional) Did anything else about this person stand out to you?

APPENDIX B

By-Speaker Social Evaluations

Category	Rating Scale	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Sp7	Sp8	Sp9	Sp10	Sp11
Ethnic Background	Asian	4.89	2.76	3.44	2.83	2.76	3.88	3.29	4.28	5.39	4.06	2.94
	White	4.17	6.00	4.56	5.94	4.94	5.82	6.47	5.11	3.22	5.44	5.89
	Latin	3.61	3.00	2.44	2.56	3.53	3.41	3.18	2.89	2.24	3.11	2.67
	Black	2.56	2.53	2.50	2.28	3.71	2.76	2.24	2.72	1.67	2.44	2.39
Regional Origin	Asia	2.89	2.33	3.06	2.17	2.00	2.78	2.28	2.67	4.61	2.56	1.89
	US/Canada	5.44	6.00	5.17	6.61	6.06	6.00	6.39	6.11	4.78	6.22	6.17
	Europe	2.94	2.89	3.83	2.83	3.17	3.50	3.22	2.33	2.67	3.00	2.72
Regional Origin (North America)	California	4.72	5.06	4.06	5.33	4.67	5.28	5.39	4.94	4.06	5.39	5.06
	Midwest	3.78	4.06	4.28	4.39	5.22	4.72	4.06	4.11	3.17	4.06	4.11
	Southern US	2.89	3.56	2.78	3.11	3.39	3.06	2.89	3.61	2.67	3.00	3.06
	elsewhere US/Canada	5.06	4.67	4.61	5.06	5.00	5.00	4.78	4.50	4.22	4.89	4.67

Table B.1: Mean ratings on demographic-related 7-point Likert scales averaged across all responses for each of the 11 voices.

Category	Rating Scale	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Sp7	Sp8	Sp9	Sp10	Sp11
Culture		5.04	6.14	4.92	6.35	5.92	6.12	6.33	6.49	4.06	5.96	6.31
	American	5.17	6.17	4.89	6.44	5.94	6.17	6.28	6.67	4.50	6.06	6.50
	Native speaker of English	5.11	6.17	5.33	6.44	6.22	6.33	6.33	6.56	3.89	6.11	6.47
	American-accented	5.00	6.06	4.83	6.28	5.89	6.00	6.33	6.56	3.89	5.89	6.22
	Aligned with American culture	4.89	6.17	4.61	6.22	5.61	6.00	6.39	6.17	3.94	5.78	6.06
Traits		3.75	4.50	3.92	4.97	3.74	5.65	5.13	4.84	4.00	5.12	4.61
	Enthusiastic	2.89	3.39	3.17	4.67	2.50	5.44	4.33	4.67	3.39	4.39	4.00
	Confident	3.72	4.89	3.78	5.67	4.35	5.78	5.28	5.83	4.56	4.33	5.44
	Friendly	4.06	4.44	4.11	4.67	3.72	5.44	5.17	4.33	3.89	5.83	4.22
	Likeable	4.11	4.61	4.17	4.67	3.83	5.78	5.28	4.56	3.78	5.83	4.67
	Attractive	3.67	4.72	3.83	4.72	3.50	6.00	5.11	4.33	3.33	5.17	4.33
	Intelligent	4.06	4.94	4.44	5.44	4.56	5.44	5.61	5.33	5.06	5.17	5.00
Style		4.76	4.76	4.78	4.56	4.61	5.07	4.94	4.49	4.39	5.18	4.69
	Feminine	4.67	5.50	5.06	5.67	4.67	6.22	5.67	5.11	4.61	6.22	5.28
	Clear speaker	4.67	5.78	5.22	6.00	5.94	6.22	5.67	6.17	5.11	5.78	5.89
	Cool	4.06	4.39	3.89	4.56	4.00	5.44	4.78	4.67	3.83	4.50	4.61
	Casual	4.94	3.61	4.11	3.61	3.61	4.61	4.28	4.00	3.72	5.50	4.56
	Nerdy	4.72	3.83	4.44	4.11	4.11	3.39	4.44	3.94	4.50	4.39	3.17
	Slow speaker	5.50	5.44	5.94	3.44	5.33	4.56	4.78	3.06	4.56	4.67	4.67

Table B.2: Mean ratings on culture-related 7-point semantic differential scales averaged across all responses for each of the 11 voices. Scale labels representing the interpretation of the maximum value of 7 (reference term).

APPENDIX C

Experimental Stimuli

Variable	Word	Variable	Word
/oʊ/-backing	coding	/u/-fronting	kudo
	goaded		moody
	hoses		noodle
	modem		rudeness
	posers		brooding
	robot		prudish
	sojourn		students
	soda		scooted
	podium		soothingly
	soberness		studious
/ð/-stopping	featherweight	/æN/-raising	candied
	leathery		channeled
	mothership		dancer
	worthiest		fancy
	otherness		tantrums
	heathers		vanish
	lathers		manager
	othered		randomness
	fathered		banditry
	weathered		sanity
PIN-PEN	dental	/æ/-backing	battles
	friendless		madden
	gemstone		padding
	lemon		saddest
	spenders		shattered
	tendons		tatters
	venture		battery
	generates		gatherings
	sentiments		tattletales
	ventilates		chattering
FEEL-FILL	fieldwork	/t/-deletion	coastguard
	ceiling		fastball
	healers		lastly
	mealtime		restful
	peeling		softball
	wheelchair		vastly
	wielding		waistline
	steelworks		restlessness
	dealership		wistfulness
	wheelbarrow		wristwatches

Table C.1: Wordlist for auditory stimuli in speaker identification experiment.

APPENDIX D

Sociolinguistic Background Questionnaire

Section 1

This section asks what you think about certain cultural ideas relating to Asian Americans.

The term "Asian American" may be defined as "an American who is of Asian descent", but "Asian American" may also be interpreted differently by different people.

What kind of people would you typically describe as 'Asian American'?

[ ]

Do you personally identify as "Asian American"?

- ❍ Yes
- ❍ No
- ❍ It depends
- ❍ Other [ ]

To what extent do you personally identify as "Asian American"?

- ❍ Not at all (1)
- ❍ Slightly (2)
- ❍ Somewhat (3)
- ❍ Moderately (4)
- ❍ Significantly (5)
- ❍ Considerably (6)
- ❍ Completely (7)

In this survey, we often ask about "Asian Americans" who grew up in North America (i.e., second and later generation immigrants) as compared to those who immigrated as an adult (i.e., first generation immigrants).

Next, we ask about your thoughts and experiences with three potential types of Asian Americans that have been described by other Asian Americans.

Important: Note that you may consider some of the terminology used here to be negative, demeaning, or derogatory. In this survey, we use them only to help describe these ideas as expressed by some Asian Americans.

Start of Loop ______________________________

Consider an Asian American who *[Persona Description]*

This survey will use the term "*[Persona Label]*" Asian American to refer to the concept described above.

How often does this concept of "*[Persona Label]*" Asian Americans come up in your daily experiences? (e.g., hearing about it in conversations, reading about it in online media, noticing it in your social interactions, thinking about it.)

- ❍ Never (1)
- ❍ Rarely (2)

Loop	*Persona Label*	*Persona Description*
1	bicultural	is American-born and/or raised and who grew up with both American culture and their Asian ethnic culture. They speak English fluently and often understand or speak their heritage language to some degree. They tend to hang out with people of a variety of backgrounds, both Asian and non-Asian. Some might describe this recognizable type of Asian American as a "normal" or "regular" Asian American.
2	culturally Asian	grew up in America but is more closely connected to Asian culture, including enjoying Asian media and wearing Asian-inspired fashion. They tend to hang out with others who speak their heritage language, and carry an accent when speaking English. Some people might describe this recognizable type of Asian American as a "FOB" or "fobby" Asian American.
3	culturally American	grew up in America and is more closely connected to American culture, like enjoying American food and wearing American-style clothing. They tend to only speak English and hang out with white or non-Asian Americans. Some people might describe this recognizable type of Asian American as a "banana" or "whitewashed" Asian American.

❍ Occasionally (3)

❍ Sometimes (4)

❍ Commonly (5)

❍ Frequently (6)

❍ All the time (7)

How familiar is this concept of "*[Persona Label]*" Asian Americans to you?

❍ Not familiar at all (1)

❍ Slightly familiar (2)

❍ Somewhat familiar (3)

❍ Moderately familiar (4)

❍ Quite familiar (5)

❍ Very familiar (6)

❍ Extremely familiar (7)

When did you become familiar with this concept of "*[Persona Label]*" Asian Americans?

- ❍ Never before now (1)
- ❍ Very recently (2)
- ❍ Somewhat recently (3)
- ❍ Some time ago (4)
- ❍ Somewhat early on (5)
- ❍ Very early on (6)
- ❍ As long as I can remember (7)

Think about the people that you encounter in your personal life (e.g., social networks, local community). How often do you feel you notice examples of "*[Persona Label]*" Asian Americans?

- ❍ Never (1)
- ❍ Rarely (2)
- ❍ Occasionally (3)
- ❍ Sometimes (4)
- ❍ Commonly (5)
- ❍ Frequently (6)
- ❍ All the time (7)

Think about the people that you encounter in media (e.g., YouTube, Instagram, TV shows, movies). How often do you feel you notice examples of "*[Persona Label]*" Asian Americans?

- ❍ Never (1)
- ❍ Rarely (2)
- ❍ Occasionally (3)
- ❍ Sometimes (4)
- ❍ Commonly (5)

- ❍ Frequently (6)
- ❍ All the time (7)

How realistic, accurate, or true do you think this concept of "*[Persona Label]*" Asian Americans is?

- ❍ Not at all true (1)
- ❍ Slightly true (2)
- ❍ Somewhat true (3)
- ❍ Moderately true (4)
- ❍ Quite true (5)
- ❍ Very true (6)
- ❍ Extremely true (7)

Do you consider this concept of "*[Persona Label]*" Asian Americans to be positive, negative, or neutral?

- ❍ Extremely negative (1)
- ❍ Negative (2)
- ❍ Slightly negative (3)
- ❍ Neutral (4)
- ❍ Slightly positive (5)
- ❍ Positive (6)
- ❍ Extremely positive (7)

To what extent do you feel like "*[Persona Label]*" Asian American describes your experiences?

- ❍ Not at all (1)
- ❍ Slightly (2)
- ❍ Somewhat (3)
- ❍ Moderately (4)

- ❍ Significantly (5)
- ❍ Considerably (6)
- ❍ Completely (7)

To what extent do you personally identify as a "*[Persona Label]*" Asian American?

- ❍ Not at all (1)
- ❍ Slightly (2)
- ❍ Somewhat (3)
- ❍ Moderately (4)
- ❍ Significantly (5)
- ❍ Considerably (6)
- ❍ Completely (7)

(Optional) Feel free to share any other thoughts about the concept of "*[Persona Label]*" Asian Americans.

End of Loop

Section 2

This section asks what you think about how Asian Americans speak.

Generally speaking, do you think that there are differences between how some Asian Americans speak and how other Americans speak?

❍ Yes

❍ No

❍ Other (Please explain) [ ]

There are people who say they can tell when somebody sounds like they are an Asian American who grew up in North America. Essentially, this is the idea that some (but not necessarily all) Asian American people "sound Asian (American)" or have an "Asian American accent" when speaking English.

This survey will use the term "Asian American accent" to refer to the concept described above.

(Optional) What about a person's voice or speaking style do you think might "sound Asian" to you or others? How would you describe an "Asian American accent" or "Asian-sounding speech"? Be as specific or as general as you wish.

[ ]

Do you think that "Asian American accented speech" and "Asian foreign accented speech" are different?

❍ Yes

❍ No

❍ Other (Please explain) [ ]

(Optional) In what ways do you think “Asian American accented speech” and “Asian foreign accented speech” might be similar or different? Be as specific or as general as you wish. For example, are they similar in quality but different in degree (e.g., one is just a “stronger accent”)?

How often does the idea of an “Asian American accent” come up in your daily experiences? (e.g., hearing about it in conversations, reading about it in online media, noticing it in your social interactions, thinking about it)

- ❍ Never (1)
- ❍ Rarely (2)
- ❍ Occasionally (3)
- ❍ Sometimes (4)
- ❍ Commonly (5)
- ❍ Frequently (6)
- ❍ All the time (7)

How familiar is the idea of an “Asian American accent” to you?

- ❍ Not familiar at all (1)
- ❍ Slightly familiar (2)
- ❍ Somewhat familiar (3)
- ❍ Moderately familiar (4)
- ❍ Quite familiar (5)
- ❍ Very familiar (6)
- ❍ Extremely familiar (7)

When did you become familiar with with this idea of an “Asian American accent”?

- ❍ Never before now (1)

- ❍ Very recently (2)
- ❍ Somewhat recently (3)
- ❍ Some time ago (4)
- ❍ Somewhat early on (5)
- ❍ Very early on (6)
- ❍ As long as I can remember (7)

Think about the people that you encounter in your personal life (e.g., social networks, local community). How often do you feel you notice examples of an "Asian American accent"?

- ❍ Never (1)
- ❍ Rarely (2)
- ❍ Occasionally (3)
- ❍ Sometimes (4)
- ❍ Commonly (5)
- ❍ Frequently (6)
- ❍ All the time (7)

Think about the people that you encounter in media (e.g., YouTube, Instagram, TV shows, movies). How often do you feel you notice examples of an "Asian American accent"?

- ❍ Never (1)
- ❍ Rarely (2)
- ❍ Occasionally (3)
- ❍ Sometimes (4)
- ❍ Commonly (5)
- ❍ Frequently (6)
- ❍ All the time (7)

How realistic, accurate, or true do you think the idea of an “Asian American accent” is?

❍ Not at all true (1)

❍ Slightly true (2)

❍ Somewhat true (3)

❍ Moderately true (4)

❍ Quite true (5)

❍ Very true (6)

❍ Extremely true (7)

Do you consider the idea of an “Asian American accent” to be positive, negative, or neutral?

❍ Extremely negative (1)

❍ Negative (2)

❍ Slightly negative (3)

❍ Neutral (4)

❍ Slightly positive (5)

❍ Positive (6)

❍ Extremely positive (7)

How often have other people told you that you “sound Asian” or speak with an “Asian American accent”?

❍ Never (1)

❍ Rarely (2)

❍ Occasionally (3)

❍ Sometimes (4)

❍ Commonly (5)

❍ Frequently (6)

❍ All the time (7)

To what extent do you feel like you might "sound Asian" or speak with an "Asian American accent"?

- ❍ Not at all (1)
- ❍ Slightly (2)
- ❍ Somewhat (3)
- ❍ Moderately (4)
- ❍ Significantly (5)
- ❍ Considerably (6)
- ❍ Completely (7)

(Optional) Feel free to share any other thoughts about the idea of an "Asian American accent".

Each person might have different ideas about what exactly "sounding Asian" is or what makes up an "Asian American accent". Plus, some people may not have come across the concept before.

Next, we ask about your thoughts and experiences with two potential aspects of an Asian American accent that have been described by other Asian Americans.

Start of Loop

Loop	*Variable Label*	*Variable Examples*	*Variable Description*
1	"o" with "more emphasis"	no, grow, home, and boat (with an "o" sound)	the "o" sound as more "drawn out", "longer", "elongated", or "emphasized". It may also be described as sounding "rounder" and "lower in tone".
2	"th" more like "d"	that, though, brother, and rather (with a "th" sound)	the "th" sound with a "d" sound.

Do you think that some Asian Americans pronounce words like [*Variable Examples*] differently from White Americans?

- ❍ Yes
- ❍ No
- ❍ Other [ ]

If you can, explain how you think the pronunciation is different for some Asian Americans.

[ ]

Consider words like [*Variable Examples*]. There are people who think that some Asian Americans pronounce [*Variable Description*]

For shorthand, this survey will use the phrase 'pronouncing [*Variable Label*]' to refer to the concept described above.

How often does the idea that some Asian Americans pronounce [*Variable Label*] come up in your daily experiences? (e.g., hearing about it in conversations, reading about it in online media, noticing it in your social interactions, thinking about it)

- ❍ Never (1)
- ❍ Rarely (2)
- ❍ Occasionally (3)
- ❍ Sometimes (4)
- ❍ Commonly (5)
- ❍ Frequently (6)
- ❍ All the time (7)

How familiar is the idea that some Asian Americans pronounce [*Variable Label*] to you?

❍ Not familiar at all (1)

❍ Slightly familiar (2)

❍ Somewhat familiar (3)

❍ Moderately familiar (4)

❍ Quite familiar (5)

❍ Very familiar (6)

❍ Extremely familiar (7)

When did you become familiar with the idea that some Asian Americans pronounce [*Variable Label*]?

❍ Never before now (1)

❍ Very recently (2)

❍ Somewhat recently (3)

❍ Some time ago (4)

❍ Somewhat early on (5)

❍ Very early on (6)

❍ As long as I can remember (7)

Think about the people that you hear speaking in your personal life (e.g., social networks, local community). How often do you feel you notice examples of Asian Americans pronouncing [*Variable Label*]?

❍ Never (1)

❍ Rarely (2)

❍ Occasionally (3)

❍ Sometimes (4)

❍ Commonly (5)

❍ Frequently (6)

❍ All the time (7)

Think about the people that you hear speaking in media (e.g., YouTube, Instagram, TV shows, movies). How often do you feel you notice examples of Asian Americans pronouncing [*Variable Label*]?

❍ Never (1)

❍ Rarely (2)

❍ Occasionally (3)

❍ Sometimes (4)

❍ Commonly (5)

❍ Frequently (6)

❍ All the time (7)

How realistic, accurate, or true do you think the idea that some Asian Americans pronounce [*Variable Label*] is?

❍ Not at all true (1)

❍ Slightly true (2)

❍ Somewhat true (3)

❍ Moderately true (4)

❍ Quite true (5)

❍ Very true (6)

❍ Extremely true (7)

Do you consider the idea that some Asian Americans pronounce [*Variable Label*] to be positive, negative, or neutral?

❍ Extremely negative (1)

❍ Negative (2)

❍ Slightly negative (3)

❍ Neutral (4)

❍ Slightly positive (5)

❍ Positive (6)

❍ Extremely positive (7)

How often have other people told you that you pronounce [*Variable Label*]?

- ❍ Never (1)
- ❍ Rarely (2)
- ❍ Occasionally (3)
- ❍ Sometimes (4)
- ❍ Commonly (5)
- ❍ Frequently (6)
- ❍ All the time (7)

To what extent do you feel like you might pronounce [*Variable Label*]?

- ❍ Not at all (1)
- ❍ Slightly (2)
- ❍ Somewhat (3)
- ❍ Moderately (4)
- ❍ Significantly (5)
- ❍ Considerably (6)
- ❍ Completely (7)

(Optional) Feel free to share any other thoughts on Asian Americans pronouncing [*Variable Label*].

End of Loop ____

BIBLIOGRAPHY

Alderton, R. (2020). Speaker Gender and Salience in Sociolinguistic Speech Perception: Goose-fronting in Standard Southern British English. *Journal of English Linguistics*, *48*(1), 72–96. https://doi.org/10.1177/0075424219896400

Ambridge, B. (2020a). Abstractions made of exemplars or 'You're all right, and I've changed my mind': Response to commentators. *First Language*, *40*(5-6), 640–659. https://doi.org/10.1177/0142723720949723

Ambridge, B. (2020b). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*(5-6), 509–559. https://doi.org/10.1177/0142723719869731

Auer, P., Barden, B., & Grosskopf, B. (1998). Subjective and Objective Parameters Determining 'Salience' in Long-term Dialect Accommodation. *Journal of Sociolinguistics*, *2*(2), 163–187. https://doi.org/10.1111/1467-9481.00039

Babel, M. (2022). Adaptation to Social-Linguistic Associations in Audio-Visual Speech. *Brain Sciences*, *12*(7), 845. https://doi.org/10.3390/brainsci12070845

Babel, M., & Mellesmoen, G. (2019). Perceptual Adaptation to Stereotyped Accents in Audio-Visual Speech (S. Calhoun, P. Escudero, M. Tabain, & P. Warren, Eds.). In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*.

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, *137*(5), 2823–2833. https://doi.org/10.1121/1.4919317

Baranowski, M. (2013). On the role of social factors in the loss of phonemic distinctions1. *English Language & Linguistics*, *17*(2), 271–295. https://doi.org/10.1017/S1360674313000038

Bargh, J. A. (1992). The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects. *The American Journal of Psychology*, *105*(2)jstor 1423027, 181. https://doi.org/10.2307/1423027

Barreda, S. (2021). Fast Track: Fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard*, *7*(1). https://doi.org/10.1515/lingvan-2020-0051

Barreda, S., & Nearey, T. M. (2018). A regression approach to vowel normalization for missing and unbalanced data. *The Journal of the Acoustical Society of America*, *144*(1), 500–520. https://doi.org/10.1121/1.5047742

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauman, C. (2013). Social Evaluation of Asian Accented English. *University of Pennsylvania Working Papers in Linguistics*, *19*(2).

Bauman, C. (2016). *Speaking of Sisterhood: A Sociolinguistic Study of an Asian American Sorority* (Doctoral dissertation). New York University.

Becker, K. (2014). The social motivations of reversal: Raised bought in New York City English. *Language in Society*, *43*(4), 395–420. https://doi.org/http://dx.doi.org/10.1017/S0047404514000372

Benor, S. B. (2010). Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity1. *Journal of Sociolinguistics*, *14*(2), 159–183. https://doi.org/10.1111/j.1467-9841.2010.00440.x

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational* *and Psychological Measurement*, *65*(5), 676–696. https://doi.org/10.1177/0013164404272507

Blazej, L. J., & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology. Human Perception and Performance*, *41*(1), 50–68. https://doi.org/10.1037/a0038509

Boersma, P., & Weenick, D. (2022). Praat: Doing Phonetics by Computer.

Bouavichith, D. A., Calloway, I., Craft, J. T., Hildebrandt, T., Tobin, S. J., & Beddor, P. S. (2019). Perceptual Influences of Social and Linguistic Priming Are Bidirectional. *Proceedings of the 19th International Congress of Phonetic Sciences*, 5.

Bucholtz, M., Bermudez, N., Fung, V., Edwards, L., & Vargas, R. (2007). Hella Nor Cal or Totally So Cal?: The Perceptual Dialectology of California. *Journal of English Linguistics*, *35*(4), 325–352. https://doi.org/10.1177/0075424207307780

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Campbell-Kibler, K. (2007). Accent, (ING), and the Social Logic of Listener Perceptions. *American Speech*, *82*(1), 32–64. https://doi.org/10.1215/00031283-2007-002

Campbell-Kibler, K. (2010a). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change*, *22*(3), 423–441. https://doi.org/10.1017/S0954394510000177

Campbell-Kibler, K. (2010b). Sociolinguistics and Perception. *Language and Linguistics Compass*, *4*(6), 377–389. https://doi.org/10.1111/j.1749-818X.2010.00201.x

Campbell-Kibler, K. (2016). Towards a cognitively realistic model of meaningful sociolinguistic variation. In A. M. Babel (Ed.), *Awareness and Control in Soci-*

olinguistic Research (pp. 123–151). Cambridge University Press. https://doi.org/10.1017/CBO9781139680448.008

Cargile, A. C. (1997). Attitudes toward Chinese-Accented Speech: An Investigation in Two Contexts. *Journal of Language and Social Psychology*, *16*(4), 434–443. https://doi.org/10.1177/0261927X970164004

Chang, Y.-H. S. (2017). The Influence of Dialect Information on the Perception of the Mandarin Alveolar-Retroflex Contrast. *Concentric: Studies in Linguistics*, *43*(1), 1–23. https://doi.org/10.6241/concentric.ling.43.1.01

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1–36.

Cheng, A., Faytak, M., & Cychosz, M. (2016). Language, race, and vowel space: Contemporary Californian English (E. Clem, V. Dawson, A. Shen, & A. Horan, Eds.). In E. Clem, V. Dawson, A. Shen, & A. Horan (Eds.), *Proceedings of the Forty-Second Annual Meeting of the Berkeley Linguistics Society*.

Cheng, A. (2020). *Accent and Ideology among Bilingual Korean Americans* (Doctoral Disseration). University of California, Berkeley.

Cheng, A. (2022). Second Dialect Acquisition "in real time": Two longitudinal case studies from YouTube. *American Speech*, 1–33. https://doi.org/10.1215/00031283-9766922

Cheng, A., & Cho, S. (2021). The Effect of Ethnicity on Identification of Korean American Speech. *Languages*, *6*(4), 186. https://doi.org/10.3390/languages6040186

Cheng, L. S. P. (2021). *Individual variability in sociophonetic (stereotype) perception: The role of sociolinguistic awareness and sociocognitive processing style* (Master's thesis). University of Michigan. Ann Arbor.

Cheng, L. S. P., & Kramer, M. A. (2021). Introducing LingTube: An open-source toolkit for linguistic analysis of YouTube data.

Chun, E. W. (2004). Ideologies of legitimate mockery: Margaret Cho's revoicings of mock Asian. *Pragmatics*, *14*(2-3), 263–289. https://doi.org/10.1075/prag.14.2-3.10chu

Clopper, C. G. (2021). Perception of Dialect Variation. In *The Handbook of Speech Perception* (pp. 333–364). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119184096.ch13

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, *32*(1), 111–140. https://doi.org/10.1016/S0095-4470(03)00009-3

Clopper, C. G., & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change*, *18*(2), 193–221. https://doi.org/10.1017/S0954394506060091

Coats, S. (2020). Articulation Rate in American English in a Corpus of YouTube Videos. *Language and Speech*, *63*(4), 799–831. https://doi.org/10.1177/0023830919894720

Coetzee, A. W. (2018). Individual and community level variation in phonetics and phonology. In R. Mesthrie & D. Bradley (Eds.), *The Dynamics of Language: Plenary and focus lectures from the 20th International Congress of Linguists* (p. 17).

Craft, J. T., Wright, K. E., Elizabeth Weissler, R., & Queen, R. M. (2020). Language and Discrimination: Generating Meaning, Perceiving Identities, and Discriminating Outcomes. *Annual Review of Linguistics*, *6*(1), null. https://doi.org/10.1146/annurev-linguistics-011718-011659

De Houwer, J. (2006). What are Implicit Measures and Why are We Using Them? In *Handbook of Implicit Cognition and Addiction* (pp. 11–28). SAGE Publications, Inc. https://doi.org/10.4135/9781412976237.n2

De Houwer, J., & Moors, A. (2010). Implicit Measures: Similarities and Differences. In *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*. Guilford Publications.

De Vogelaer, G., Chevrot, J.-P., Katerbow, M., & Nardy, A. (2017). Chapter 1. Bridging the gap between language acquisition and sociolinguistics: Introduction to an interdisciplinary topic. In *Acquiring Sociolinguistic Variation* (pp. 1–41). https://doi.org/10.1075/silv.20.01dev

Dehaene, S. (2011). Conscious and Nonconscious Processes:Distinct Forms of Evidence Accumulation? In V. Rivasseau (Ed.), *Biological Physics: Poincaré Seminar 2009* (pp. 141–168). Springer. https://doi.org/10.1007/978-3-0346-0428-4_7

Dehaene, S., Changeux, J.-P., & Joliot, F. (2003). Neural Mechanisms for Access to Consciousness, 37.

Devos, T., & Banaji, M. R. (2005). American = White? *Journal of Personality and Social Psychology*, *88*(3), 447–466. https://doi.org/10.1037/0022-3514.88.3.447

Docherty, G. J., & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*, *142*, 42–56. https://doi.org/10.1016/j.lingua.2013.01.011

Docherty, G. J., Langstrof, C., & Foulkes, P. (2013). Listener evaluation of sociophonetic variability: Probing constraints and capabilities. *Linguistics*, *51*(2), 355–380. https://doi.org/10.1515/ling-2013-0014

D'Onofrio, A. (2018a). Controlled and automatic perceptions of a sociolinguistic marker. *Language Variation and Change*, *30*(2), 261–285. https://doi.org/10.1017/S095439451800008X

D'Onofrio, A. (2018b). Personae and phonetic detail in sociolinguistic signs. *Language in Society*, *47*(4), 513–539. https://doi.org/10.1017/S0047404518000581

D'Onofrio, A. (2019). Complicating categories: Personae mediate racialized expectations of non-native speech. *Journal of Sociolinguistics*, *23*(4), 346–366. https://doi.org/10.1111/josl.12368

D'Onofrio, A. (2020). Personae in sociolinguistic variation. *WIREs Cognitive Science*, *11*(6), e1543. https://doi.org/10.1002/wcs.1543

D'Onofrio, A. (2021a). Age-based perceptions of a reversing regional sound change. *Journal of Phonetics*, *86*, 101038. https://doi.org/10.1016/j.wocn.2021.101038

D'Onofrio, A. (2021b). Sociolinguistic Signs as Cognitive Representations. In E. Moore, L. Hall-Lew, & R. J. Podesva (Eds.), *Social Meaning and Linguistic Variation: Theorizing the Third Wave* (pp. 153–175). Cambridge University Press. https://doi.org/10.1017/9781108578684.007

D'Onofrio, A., & Benheim, J. (2020). Contextualizing reversal: Local dynamics of the Northern Cities Shift in a Chicago community. *Journal of Sociolinguistics*, *24*(4), 469–491. https://doi.org/10.1111/josl.12398

D'Onofrio, A., & van Hofwegen, J. (2020). Nisei Style: Vowel Dynamism in a Second-Generation Japanese American Community. *The Publication of the American Dialect Society*, *105*(1), 79–94. https://doi.org/10.1215/00031283-8820631

Drager, K. (2010). Sociophonetic Variation in Speech Perception. *Language and Linguistics Compass*, *4*(7), 473–480. https://doi.org/10.1111/j.1749-818X.2010.00210.x

Drager, K. (2011). Speaker Age and Vowel Perception. *Language and Speech*, *54*(1), 99–121. https://doi.org/10.1177/0023830910388017

Drager, K., Hardeman-Guthrie, K., Schutz, R., & Chik, I. (2021). Perceptions of Style: A Focus on Fundamental Frequency and Perceived Social Characteristics. In E. Moore, L. Hall-Lew, & R. J. Podesva (Eds.), *Social Meaning and Linguistic Variation: Theorizing the Third Wave* (pp. 176–202). Cambridge University Press. https://doi.org/10.1017/9781108578684.008

Drager, K., & Kirtley, M. J. (2016). Awareness, Salience, and Stereotypes in Exemplar-Based Models of Speech Production and Perception. In A. M. Babel (Ed.), *Awareness and Control in Sociolinguistic Research* (pp. 1–24). Cambridge University Press. https://doi.org/10.1017/CBO9781139680448.003

Eckert, P. (2008a). Variation and the indexical field. *Journal of Sociolinguistics*, *12*(4), 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

Eckert, P. (2008b). Where do ethnolects stop? *International Journal of Bilingualism*, *12*(1-2), 25–42. https://doi.org/10.1177/13670069080120010301

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, *41*(1), 87–100. https://doi.org/10.1146/annurev-anthro-092611-145828

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology*, *1*(1), 5–39. https://doi.org/10.1515/labphon.2010.003

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438. https://doi.org/10.1016/j.wocn.2005.08.002

Freeman, J. B., & Johnson, K. L. (2016). More Than Meets the Eye: Split-Second Social Perception. *Trends in Cognitive Sciences*, *20*(5), 362–374. https://doi.org/10.1016/j.tics.2016.03.003

Fridland, V., & Bartlett, K. (2006). The social and linguistic conditioning of back vowel fronting across ethnic groups in Memphis, Tennessee. *En-*

glish Language & Linguistics, *10*(1), 1–22. https : / / doi . org / 10 . 1017 / S1360674305001681

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515–546). De Gruyter Mouton.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Günther, F., Müller, H. J., & Geyer, T. (2017). Salience, attention, and perception. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 289–312). De Gruyter Mouton. https://doi.org/10.1037/15969-014

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (Eighth edition). Cengage.

Hall-Lew, L. (2009). *Ethnicity and phonetic variation in a San Francisco neighborhood* (Doctoral dissertation). Stanford University. United States – California.

Hall-Lew, L., & Boyd, Z. (2020). Sociophonetic perspectives on stylistic diversity in speech research. *Linguistics Vanguard*, *6*(s1). https://doi.org/10.1515/lingvan-2018-0063

Hanna, D. B. (1997). Do I Sound "Asian" to You?: Linguistic Markers of Asian American Identity. *University of Pennsylvania Working Papers in Linguistics*, *4*(2), 15.

Hay, J. (2018). Sociophonetics: The Role of Words, the Role of Context, and the Role of Words in Context. *Topics in Cognitive Science*, *10*(4), 696–706. https://doi.org/10.1111/tops.12326

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 865–892. https://doi.org/10.1515/ling.2010.027

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3), 351–379. https://doi.org/10.1515/TLR.2006.014

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

Hickey, R. (2000). Salience, stigma and standard. In L. Wright & J. Algeo (Eds.), *The Development of Standard English, 1300-1800: Theories, Descriptions, Conflicts*. Cambridge University Press.

Hosoda, M., Stone-Romero, E. F., & Walter, J. N. (2007). Listeners' Cognitive and Affective Reactions to English Speakers with Standard American English and Asian Accents. *Perceptual and Motor Skills*, *104*(1), 307–326. https://doi.org/10.2466/pms.104.1.307-326

Irvine, J. T., & Gal, S. (2000). Language ideology and linguistic differentiation. In P. V. Kroskrity (Ed.), *Regimes of language: Ideologies, polities, and identities* (pp. 35–84). School of American Research Press.

Jensen, E., Jones, N., Orozco, K., Medina, L., Perry, M., Bolender, B., & Battle, K. (2021). Measuring Racial and Ethnic Diversity for the 2020 Census.

Jensen, M. M. (2016). Linking Place and Mind: Localness As a Factor in Socio-Cognitive Salience. *Frontiers in Psychology*, *7*.

Jeon, L. (2017). Korean Ethnic Orientation and Regional Linguistic Variability in the Multiethnic Context of Houston. In C. J. Jenks & J. W. Lee (Eds.), *Korean Englishes in Transnational Contexts* (pp. 93–114). Springer International Publishing. https://doi.org/10.1007/978-3-319-59788-1_5

Jeon, M. (2007). Language Ideologies and Bilingual Education: A Korean-American Perspective. *Language Awareness*, *16*(2), 114–130. https://doi.org/10.2167/la369.0

Johnson, K., Johnson, K., & Mullennix, J. (1997). Speech Perception without Speaker Normalization: An Exemplar Model. In *Talker Variability in Speech Processing*. (pp. 145–165.). Academic Press.

Jones, J. A., & Renwick, M. E. L. (2021). Spatial analysis of sub-regional variation in Southern US English. *Journal of Linguistic Geography*, *9*(2), 86–105. https://doi.org/10.1017/jlg.2021.7

Jones, M. N. (2019). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience*, *34*(10), 1338–1346. https://doi.org/10.1080/23273798.2018.1431679

Juskan, M. (2018). *Sound change, priming, salience : Producing and perceiving variation in Liverpool English*. Language Science Press. https://doi.org/10.5281/zenodo.1451308

Accepted: 2019-01-04 23:55

Kerswill, P., & Williams, A. (2002). "Salience" as an explanatory factor in language change: Evidence from dialect levelling in urban England. In M. C. Jones & E. Esch (Eds.), *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*. Walter de Gruyter.

Kibria, N. (1998). The contested meanings of 'Asian American': Racial dilemmas in the contemporary US. *Ethnic and Racial Studies*, *21*(5), 939–958. https://doi.org/10.1080/014198798329739

Kim, D.-E. (2021). Korean Americans in GA: [+back] on the B/OW/T. In *New Ways of Analyzing Variation 49*.

Kim, J. (2016). Perceptual Associations Between Words and Speaker Age. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1), 18. https://doi.org/10.5334/labphon.33

Kim, J., & Drager, K. (2018). Rapid Influence of Word-Talker Associations on Lexical Access. *Topics in Cognitive Science*, *10*(4), 775–786. https://doi.org/10.1111/tops.12351

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics*, *14*(2).

Kramer, A. (2021). Dependency Lengths in Speech and Writing: A Cross-Linguistic Comparison via YouDePP, a Pipeline for Scraping and Parsing YouTube Captions. *Proceedings of the Society for Computation in Linguistics*, *4*(1), 359–365.

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Mouton de Gruyter. https://doi.org/10.1515/9783110206838

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language*, *89*(1)jstor 23357721, 30–65.

Lai, W., Rácz, P., & Roberts, G. (2020). Experience With a Linguistic Variant Affects the Acquisition of Its Sociolinguistic Meaning: An Alien-Language-Learning Experiment. *Cognitive Science*, *44*(4), e12832. https://doi.org/10.1111/cogs.12832

Lawrence, D. (2015). Limited Evidence for Social Priming in the Perception of the Bath and Strut Vowels (The Scottish Consortium for ICPhS 2015, Ed.). In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, *25*, 1–18. https://doi.org/10.18637/jss.v025.i01

Lee, J., & Ramakrishnan, K. (2020). Who counts as Asian. *Ethnic and Racial Studies*, *43*(10), 1733–1756. https://doi.org/10.1080/01419870.2019.1671600

Lee, J. (2019). Many dimensions of Asian American pan-ethnicity. *Sociology Compass*, *13*(12), e12751. https://doi.org/10.1111/soc4.12751

Lee, S. (2017). Style-Shifting in Vlogging: An Acoustic Analysis of "YouTube Voice". *Lifespans and Styles*, *3*(1), 28–39. https://doi.org/10.2218/ls.v3i1.2017.1826

Leigh, D. D. (2021). *Style in Time: Online Perception of Sociolinguistic Cues* (Doctoral dissertation). Stanford University. United States – California.

Lenth, R. V. (2023). Emmeans: Estimated Marginal Means, aka Least-Squares Means.

Li, A., & Roberts, G. (2021). The emergence of indexicality in an artificial language. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Lien, P.-t., Conway, M. M., & Wong, J. (2003). The Contours and Sources of Ethnic Identity Choices Among Asian Americans. *Social Science Quarterly*, *84*(2), 461–481. https://doi.org/10.1111/1540-6237.8402015

Lin, Y.-C., Lin, P.-Y., & Yeh, L.-H. (2022). Syllable or phoneme? A mouse-tracking investigation of phonological units in Mandarin Chinese and English spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001128

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics*, *7*(3), 348–364. https://doi.org/10.1111/1467-9481.00228

Llamas, C., Watt, D., & MacFarlane, A. E. (2016). Estimating the Relative Sociolinguistic Salience of Segmental Variables in a Dialect Boundary Zone. *Frontiers in Psychology*, *7*.

Loudermilk, B. C. (2013). *Cognitive Mechanisms in the Perception of Sociolinguistic Variation* (Doctoral dissertation). University of California, Davis. United States – California.

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. https://doi.org/10.1016/j.lingua.2012.01.007

MacLeod, B. (2015). A critical evaluation of two approaches to defining perceptual salience. *Ampersand*, *2*, 83–92. https://doi.org/10.1016/j.amper.2015.07.001

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *INTERSPEECH*. https://doi.org/10.21437/interspeech.2017-1386

McGowan, K. B. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech*, *58*(4), 502–521. https://doi.org/10.1177/0023830914565191

McGowan, K. B., & Babel, A. M. (2019). Perceiving isn't believing: Divergence in levels of sociolinguistic awareness. *Language in Society*, 1–26. https://doi.org/10.1017/S0047404519000782

Montgomery, C., & Moore, E. (2018). Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech. *Language*, *94*(3), 629–661. https://doi.org/10.1353/lan.2018.0038

Nagy, N., Hoffman, M., & Walker, J. A. (2020). How do Torontonians hear ethnic identity? *Toronto Working Papers in Linguistics*, *42*. https://doi.org/10.33137/twpl.v42i1.33190

Newman, M., & Wu, A. (2011). ''Do You Sound Asian When You Speak English?" Racial Identification and Voice in Chinese and Korean Americans' English. *American Speech*, *86*(2), 152–178. https://doi.org/10.1215/00031283-1336992

Nguyen, E. (2019). *Linguistic Variation and Social Practice in a Vietnamese American Student Organization* (Doctoral dissertation). New York University. United States – New York.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85. https://doi.org/10.1177/0261927X99018001005

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, (115), 39–57.

Okamoto, D., & Mora, G. C. (2014). Panethnicity. *Annual Review of Sociology*, *40*(1), 219–239. https://doi.org/10.1146/annurev-soc-071913-043201

Park, J. (2008). Second-Generation Asian American Pan-Ethnic Identity: Pluralized Meanings of a Racial Label. *Sociological Perspectives*, *51*, 541–561. https://doi.org/10.1525/sop.2008.51.3.541

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Typological Studies in Language* (p. 137). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.08pie

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*(4), 516–530. https://doi.org/10.1016/j.wocn.2006.06.003

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*, 33–52.

Preston, D. R. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Language Awareness*, *5*(1), 40–74. https://doi.org/10.1080/09658416.1996.9959890

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. https://doi.org/10.1177/0261927X99018001002

Pyke, K., & Dang, T. (2003). ''FOB" and "Whitewashed": Identity and Internalized Racism Among Second Generation Asian Americans. *Qualitative Sociology*, *26*(2), 147–172. https://doi.org/10.1023/A:1022957011866

R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Rácz, P. (2013). *Salience in Sociolinguistics: A Quantitative Approach*. Walter de Gruyter.

Rácz, P., Beckner, C., Hay, J. B., & Pierrehumbert, J. B. (2020). Morphological convergence as on-line lexical analogy. *Language*, *96*(4), 735–770. https://doi.org/10.1353/lan.2020.0061

Rácz, P., Hay, J. B., & Pierrehumbert, J. B. (2017). Social Salience Discriminates Learnability of Contextual Cues in an Artificial Language. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00051

Renn, K. A. (2012). Creating and Re-Creating Race: The Emergence of Racial Identity as a Critical Element in Psychological, Sociological and Ecological Perspectives on Human Development. In C. Wijeyesinghe & B. W. Jackson

(Eds.), *New perspectives on racial identity development: Integrating emerging frameworks* (2nd ed). New York University Press.

Reyes, A., & Lo, A. (Eds.). (2009). *Beyond Yellow English: Toward a Linguistic Anthropology of Asian Pacific America*. Oxford University Press, USA.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531. https://doi.org/10.1007/BF00973770

Sanker, C., Babinski, S., Burns, R., Evans, M., Johns, J., Kim, J., Smith, S., Weber, N., & Bowern, C. (2021). (Don't) try this at home! The effects of recording devices and software on phonetic analysis. *Language*, *97*(4), e360–e382.

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2011). You had me at "Hello": Rapid extraction of dialect information from spoken words. *NeuroImage*, *56*(4), 2329–2338. https://doi.org/10.1016/j.neuroimage.2011.04.007

Schirra, R. E. (2012). *Attitudes Toward Korean-Accented and Korean American English* (Thesis)
Accepted: 2012-09-13T17:34:30Z.

Schmid, H.-J., & Günther, F. (2016). Toward a Unified Socio-Cognitive Framework for Salience in Language. *Frontiers in Psychology*, *7*.

Schneider, D. J. (2005). *The Psychology of Stereotyping*. Guilford Press.

Schneider, E. W. (2016). World Englishes on YouTube: Treasure trove or nightmare? *World Englishes*, 253–282.

Sewell, A., & Chan, J. (2010). Patterns of variation in the consonantal phonology of Hong Kong English. *English World-Wide*, *31*(2), 138–161. https://doi.org/10.1075/eww.31.2.02sew

Shankar, S. (2008). Speaking like a Model Minority: ''FOB'' Styles, Gender, and Racial Meanings among Desi Teens in Silicon Valley. *Journal of Linguistic*

Anthropology, *18*(2), 268–289. https://doi.org/10.1111/j.1548-1395.2008.00022.x

Sheydaei Baghdadeh, I. (2021). *Local and Ethnic Identities: MENA-Americans' Linguistic Behavior and Ethnic Rootedness* (Doctoral dissertation). The University of Wisconsin - Madison. United States – Wisconsin.

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), 16–26. https://doi.org/10.1037/0003-066X.60.1.16

Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*(1), 3–21. https://doi.org/10.1037/0033-295X.99.1.3

Sneller, B., & Roberts, G. (2018). Why some behaviors spread while others don't: A laboratory simulation of dialect contact. *Cognition*, *170*, 298–311. https://doi.org/10.1016/j.cognition.2017.10.014

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Staum Casasanto, L. (2010). What do Listeners Know about Sociolinguistic Variation? *University of Pennsylvania Working Papers in Linguistics*, *15*(2).

Stevens, G., Ishizawa, H., & Grbic, D. (2015). Measuring race and ethnicity in the censuses of Australia, Canada, and the United States: Parallels and paradoxes. *Canadian Studies in Population [ARCHIVES]*, *42*(1-2), 13–34. https://doi.org/10.25336/P6PW39

Strand, E. A. (1999). Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology*, *18*(1), 86–100. https://doi.org/10.1177/0261927X99018001006

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in psychology*, *4*, 1015. https://doi.org/10.3389/fpsyg.2013.01015

Thomas, E. R. (2007). Phonological and Phonetic Characteristics of African American Vernacular English. *Language and Linguistics Compass*, *1*(5), 450–475. https://doi.org/10.1111/j.1749-818X.2007.00029.x

Thomas, E. R., & Carter, P. M. (2006). Prosodic rhythm and African American English. *English World-Wide*, *27*(3), 331–355. https://doi.org/10.1075/eww.27.3.06tho

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. https://doi.org/10.1111/j.1467-9841.2004.00251.x

Trudgill, P. (1986). *Dialects in contact*. Blackwell.

Tuan, M. (1998). *Forever Foreigners Or Honorary Whites?: The Asian Ethnic Experience Today*. Rutgers University Press.

Walker, M., Szakay, A., & Cox, F. (2019). Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners? *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 7. https://doi.org/10.5334/labphon.90

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522. https://doi.org/10.1016/j.wocn.2007.02.003

Wolfram, W., & Schilling, N. (2015). *American English: Dialects and Variation*. John Wiley & Sons, Incorporated.

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

Wong, P., & Babel, M. (2017). Perceptual identification of talker ethnicity in Vancouver English. *Journal of Sociolinguistics*, *21*(5), 603–628. https://doi.org/10.1111/josl.12264

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Woolard, K. A., & Schieffelin, B. B. (1994). Language Ideology. *Annual Review of Anthropology*, *23*, 55–82.

Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody*.

Yao, Y., & Chang, C. B. (2016). On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese. *Language*, *92*(2), 433–467. https://doi.org/10.1353/lan.2016.0031

Yi, H.-G., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America*, *134*(5), EL387–EL393. https://doi.org/10.1121/1.4822320

Yu, A. C. L., & Zellou, G. (2019). Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics*, *5*(1), 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815

Zarcone, A., van Schijndel, M., Vogels, J., & Demberg, V. (2016). Salience and Attention in Surprisal-Based Accounts of Language Processing. *Frontiers in Psychology*, *7*.

Zheng, M. (2018). *You Have to Learn to Adapt: A Sociolinguistic Study of Chinese Americans in the "Asian City" of Southeast Michigan* (Doctoral dissertation). Michigan State University. United States – Michigan.

Zheng, Y., & Samuel, A. G. (2017). Does seeing an Asian face make speech sound more accented? *Attention, Perception, & Psychophysics*, *79*(6), 1841–1859. https://doi.org/10.3758/s13414-017-1329-2

Zipp, L., & Staicov, A. (2016). English in San Francisco Chinatown: Indexing identity with speech rhythm? *World Englishes*, 205–228.