# Unpacking Listener Bias:
# The Effects of Attitudes, Knowledge, and Experience on Clinician's Ratings of African-American Children's Speech

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Karen Elizabeth Evans

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Advisor: Benjamin Munson, Ph.D.

July 2012

**Acknowledgements**

There are many people to whom I owe thanks and credit for this project. First and foremost, I am grateful for the invaluable guidance and support of my advisor, Dr. Ben Munson, through this entire process. To the other members of the "thesis club," especially Amy Andrzejewski, I couldn't have made it through without you. To all the other labbies, particularly the supremely competent Mandi Proue, thank you for all your help with running subjects. To Dr. Jan Edwards, thank you for your generosity in sharing the WID stimuli with me; it has been an amazing resource without which this project could not have happened. I also owe an intellectual debt to Andrea Christy and Molly Babel, whose stimuli and protocols are used throughout. Finally, I would like to thank all my friends and family for their love and support.

## Dedication

To the membership of the DSS.

## Table of Contents


Introduction ..... 1

Overview ..... 1

African American English ..... 2

Effects of Context and Talker Characteristics on Perception ..... 3

Potential Sources of Clinician Bias ..... 8

The Present Study and Predictions ..... 9

Methods ..... 11

Overview ..... 11

Participants ..... 11

Speech Rating Task ..... 12

Implicit Association Task ..... 17

Participant Questionnaire ..... 20

Debriefing ..... 22

Results ..... 23

Knowledge Measures ..... 23

Attitude Measures ..... 24

Speech Rating Task ..... 29

Analysis ..... 29

Discussion ..... 37

References ..... 42

Appendix A: Implicit Attitudes Test Stimuli ..... 47

Appendix B: AAE Knowledge Quiz ..... 48

Appendix C: Dialect Attitudes Survey ..... 51

Appendix D: Participant Debriefing ..... 53

# Introduction

## Overview

As the United States continues to grow in its cultural and linguistic diversity, no population reflects these demographic shifts so dramatically as that of young children. Now, for the first time, more than half of children under one year of age in the United States belong to minority groups (US Census Bureau, 2012). Many of these children will arrive at school speaking languages other than English, while others will be speakers of non-mainstream dialects such as African American English or Spanish-Influenced English. For Speech-Language Pathologists (SLPs), providing appropriate speech and language services for children from diverse backgrounds, while differentiating speech and language differences from disorders, is an ongoing challenge. Meeting this challenge requires specialized training, including knowledge of the features of various languages and dialects and alternative assessment procedures.

Another critical, though less tangible, consideration for these SLPs is the effect their own biases may have on their clinical decision-making. In its official statement on the "Knowledge and Skills Needed by Speech-Language Pathologists and Audiologists to Provide Culturally and Linguistically Appropriate Services," the American Speech-Language-Hearing Association lists knowledge of the "influence of one's own beliefs and biases in providing effective services" first among the many requisites (2004:1.1). However, the extent of this influence is not yet well understood, especially when one considers the particular influence that listener bias may exert on speech perception.

Skilled speech perception is a core competency for SLPs serving clients with phonological and articulation disorders. Articulation assessments such as the *Goldman-Fristoe Test of Articulation-Second Edition* (Goldman & Fristoe, 2000) and the *Photo Articulation Test - Third Edition* (Pendergast, Dickey, Selmar, & Soder, 1997) are scored according to subjective, binary correct-or-incorrect judgments by the examiner. SLPs use these assessments regularly to screen for and diagnose speech sound disorders in children and to determine eligibility for clinical services. If conscious or unconscious listener biases affect even a fraction of test items, the result – and its chain of consequences – could alter.

The present study is an addition to the small, but growing literature on the factors affecting adults' perception of the speech of children, particularly as related to social dialects and the presumed race of the child. The focus is on African American English (AAE), a dialect of English observed throughout the United States, and one that historically has been subject to significant cultural controversy. The following sections provide background information on AAE and review the existing literature on how talker characteristics and listener bias affect speech perception.

**African American English**

African American English is a distinct dialect of English characterized by phonological, morphological, and syntactic differences from "standard" or "mainstream" American English, although it shares many features with Southern varieties of English (see, e.g., Rickford & Rickford, 1999, for a review). It is spoken chiefly, but not only, by African Americans, and not all African Americans speak AAE. This leads to a complex relationship between individuals' attitudes toward race and their attitudes towards AAE.

Craig, Thompson, Washington and Potter (2003) compiled a taxonomy of AAE features present in the speech of elementary-aged African American children in Michigan. Some of the most salient phonological features included deletion of post-vocalic consonants (e.g. /maʊ/ for "mouth"), substitutions of /f/, /t/, /d/, or /v/ for /θ, devoicing of final consonants (e.g. /hɪs/ for "his"), and reduction of consonant clusters (e.g. /wɝl/ for "world"). See Craig et al. (1993:626-7) for a thorough chart of phonological and morphosyntactic features of AAE as spoken by children. According to the US Department of Education (2001), African American students are consistently overrepresented on special education caseloads, so the issue of adults' perception of AAE in children is critical. The next section reviews the existing literature on the way factors talkers' indexical traits, e.g., race, can shift listeners' perceptions.

**Effects of Context and Talker Characteristics on Perception**

Speech perception is a very complex skill, characterized by the fundamental lack of "invariance" between acoustic and perceptual information (Perkall & Klatt, 1986). A growing body of research shows the myriad ways that listeners overcome the invariance problem to disambiguate acoustic information and establish a steady percept. These include implicit knowledge of coarticulation (Lindblom, 1990), the effects of speaking rate (see Miller, 1981, for a review), and emotional tone of voice (Nygaard & Lunders, 2002). Another important source of information is the visual context. The classic McGurk effect (McGurk & McDonald, 1976) demonstrates that the articulatory gestures a listener sees will alter what phoneme he or she perceives, even given prior knowledge of the stimulus (Walker, Bruce, & O'Malley, 1995). Work by Tanenhaus, Spivey-Knowlton,

Eberhard, and Sedivy (1995) also shows the on-line influence of visual context in word recognition as well as syntactic parsing.

Characteristics of the talker also interact with perception in both the bottom-up and top-down directions. In the former case, listeners often form assumptions about an individual based on speech patterns, such as when one infers a person's gender and approximate age over the phone. In the latter situation, known or assumed information about the talker leads to shifts or corrections in perception. This is illustrated by an adult's ability to understand young children with developmental phonological errors (e.g., substitution of /w/ for /r/ as in /wæbɪt/). A growing body of research has examined these relationships with respect to various indexical characteristics of the talker, including age, gender, race, and region of origin, among others.

Research by Drager (2005) investigated vowel perception in a case where vowel pronunciation differs by generation. In New Zealand, younger speakers raise the vowel in "trap" to be closer to the vowel in "dress", while older speakers have similar vowels to those in American English. Drager's listeners judged more of the vowels paired with an apparently younger speaker as the raised, "dress" vowel, although the vowels were acoustically no different than those paired with apparently older speakers.

Although the New Zealand work focused on adult speakers and listeners, two studies by Munson, Edwards, Schellinger, and colleagues have found similar effects of presumed age upon perception in investigations of child speech (Munson, Edwards, Schellinger, Beckman, & Meyer, 2010; Schellinger, Edwards, Munson, & Beckman, 2008). In these studies, adult listeners rated the accuracy of productions of /s/ ranging from correct /s/ to

misarticulated /θ/. When the listeners believed the children were older, they rated productions of /θ/ as more accurate than if they believed the speakers were younger.

Known or presumed speaker gender has also been shown to affect various types of listener judgments. Strand and Johnson (1996) found when they paired phonemes with acoustic properties midway between /s/ and /θ/ with pictures of stereotypically female talkers, listeners reported perceiving /θ/ more often than when they were paired with other pictures. This suggests that listeners use their implicit knowledge of differences in sibilant production across genders, as women generally have a higher-frequency acoustic boundary between the two phonemes. Johnson, Strand, and D'Imperio (1999) later found a similar result in the perception of vowels. Listeners do not only use perceived gender to disambiguate spectral information; in a study by Munson and Seppanen (2009), adult listeners judged narratives paired with pictures of boys as being of higher quality than identical narratives paired with pictures of girls. This result suggests that listeners have differentiated expectations for the oral language of girls and boys, based on the impression that girls' skills in this area usually develop more quickly than boys'.

Speech perception also intersects with stereotypes related to cultural identity in various ways. In her dissertation research, Campbell-Kibler (2006) found that listeners formed assumptions about a speaker's intelligence based on the use of *–in'* vs. *–ing*. Niedzielski (1999) conducted an important study in which she asked residents of Detroit to match from a selection the vowels they perceived in the speech of another Detroit resident, whom they were lead to believe was either from Detroit or from Canada. The speaker's dialect, typical of middle-class residents of Detroit, included the raised-diphthong vowels seen in the Northern Cities Chain Shift dialect (NCCS). Listeners who

believed the speaker was Canadian correctly identified the raised-diphthong vowels, but listeners who believed the speaker was a fellow Detroiter perceived the diphthongs as falling closer to Mainstream American English un-raised forms. These findings support the notion that social information and the expectations it raises affect speech perception. Furthermore, the fact that Detroit residents perceived the speech of a member of their community as closer to standard than that of an outsider suggests that attitudes about non-standard dialects are also at work in perception.

Work by Hay, Drager, and colleagues (Hay, Nolan, & Drager, 2006; Drager & Hay, 2010) also examined the effects on perception of stereotypes pertaining to national origin in Australia and New Zealand. Listeners from New Zealand, primed to believe a speaker was from either Australia or New Zealand by the header (e.g., "Australian") on their response sheets, shifted their perception of the high-front lax vowel /I/ (as in "bit," which is pronounced very distinctly in those two countries) depending on the priming condition. The researchers subsequently found that the addition of an iconic national animal (a kangaroo or kiwi bird) in the testing space provided enough of a contextual clue to speaker nationality to cause this perception shift. A critical finding of this research is that these perception shifts occurred even when the participants *knew the speaker was in fact a New Zealander*. This supports the idea that the influence of unconscious factors persists despite the listener's potentially contradictory explicit knowledge.

Knowledge or assumptions about a speaker's race or ethnicity, as well as related listener attitudes, also come to bear on the listener's perceptions. Participants in Rubin's 1992 study of foreign-accented English as spoken by university teaching assistants rated the speech of the same native American English speaker as sounding more "foreign" or

non-standard when it was paired with an Asian face than when it was paired with a Caucasian face.

More recently, Staum Casasanto (2008) designed a response-time task to investigate how listeners use presumed race and knowledge of social dialects to disambiguate sentences. The stimuli in this study consisted of words like "mass," which would be indistinguishable from "mast" when produced by a speaker of AAE, in which final consonants are often reduced. Listeners were presented with these words embedded in a phrase such as "The [mæs] probably lasted…" in which the stimulus could be either "mass" or "mast" depending on context. The carrier phrases were paired with pictures of either African Americans or Caucasians. Afterwards, listeners read a phrase that completed the sentence, and response time was measured as they judged whether or not the sentence made sense. Some completed sentences made sense if the stimulus word had a reduced final cluster, as in "The [mæs] probably lasted through the storm," while others made sense only without a reduced cluster, as in "The [mæs] probably lasted an hour on Sunday." Response times differed significantly depending on the race of the picture listeners saw and the word form that correctly completes the sentence. If listeners saw a Caucasian's face, they were quicker in responding to sentences with no necessary cluster reduction, while if they saw an African American's face, they were faster in their responses to sentences where cluster reduction would be needed. This result also highlights the conflation of race with dialect in the minds of many listeners; seeing an African American face led listeners to associations with AAE, although not all African American speak that variety of English. Despite these suggestive findings, the role of

individual listener factors in mediating listener bias is still unclear. The following section addresses this question in greater detail.

**Potential Sources of Clinician Bias**

The current study is part of a larger body of research on the ways speech-language clinicians differ from untrained listeners in their speech perception, as well as the influences that beliefs, knowledge, and attitudes exert on those perceptions. This is quite a young area of exploration, but some previous work has posed variations on this question.

When discussing listeners' attitudes in relation to speech perception, it is important to distinguish *implicit* from *explicit* attitudes. There are two major problems with relying solely on measures of explicit attitudes; first, respondents are not always truthful when asked about their personal beliefs, particularly when an honest response would carry a social cost. Second, as was demonstrated in the Hay et al. (2006) study, unconscious factors can affect perception even in the face of conflicting explicit information. Fortunately, research methodologies have been developed to assess implicit, unconscious associations between concepts. Babel's (2009) dissertation examined the effects of implicit racial attitudes on speech *production*, finding that these attitudes mediated spontaneous phonetic imitation. She measured implicit attitudes using a version of the Implicit Association Task (Greenwald, McGhee, & Schwartz, 1998), an experimental paradigm based on differential reaction time that is used to measure unconscious biases.

Christy and Munson (2012) have previously examined listeners' perception of the accuracy of child speech with respect to implied speaker race, implicit racial biases, and the listeners' amount of experience perceiving child speech. They asked a group of

speech-language clinicians and a group of untrained listeners to rate acoustically modified, single-word child utterances on a continuum of articulatory accuracy. The listeners were lead to believe that the speakers were either African American or Caucasian by photographs that were presented simultaneously with the speech samples. Their stimuli consisted of /s/-initial words that were manipulated such that the /s/ was either obviously correct or clearly misarticulated as /θ/, and /t/-final words that were either correct or modified to simulate glottalization of the /t/. The simulated misarticulation in the first set of words is common of typically developing children, while glottalized final /t/ is characteristic of African American English. They found that their listeners rated words as less accurate when paired with faces of African American children versus Caucasian children. They further found this effect was mediated both by listeners' experience with children's speech and by their implicit racial attitudes. Individuals with more experience judged samples paired with African American children as more accurate than those paired with Caucasian children. Listeners whose scores on an Implicit Association Test were consistent with a pro-white bias judged the /t/-final words to be more accurate when matched with a Caucasian child, while responses of listeners with a more race-neutral IAT score patterned in the opposite direction.

**The Present Study and Predictions**

The goal of the present study was to investigate adults' judgments of the speech of children who speak African American English, including the effects of presumed speaker race, the differences between judgments of speech-language clinicians and lay listeners, and how these factors interact with listeners' knowledge of and attitudes toward AAE and African Americans. The aim was to complement and expand on previous work,

particularly that of Christy and Munson, in several ways. As in that study, an audiovisual perception experiment was conducted in which clinicians and untrained listeners rated the accuracy of words spoken by children they believed to be either African American or Caucasian. For the current study, however, the stimuli were speech samples from a large number of young AAE-speaking children of both sexes. Since these samples varied widely in articulatory accuracy and the presence of phonological dialect features, it was not necessary to simulate consonant deletions or distortions through acoustic manipulation. This change in protocol allowed investigation of whether the patterns found by Christy and Munson would be replicated and generalized in the presence of naturally varying speech accuracy and productions from genuine dialect speakers. The prediction was that the listeners would judge words to be less accurate overall when they were lead to believe the children who produced them were African American versus Caucasian. Furthermore, it was predicted that this effect would differ across listener groups.

This study is innovative in that it included not only a measure of listeners' implicit racial attitudes, but also an objective measure of their knowledge of the features of AAE and a measure of their explicit attitudes about non-mainstream dialects in general and AAE in particular. It was predicted that listeners with less knowledge of AAE would judge words to be less accurate when they believed the speaker to be African American versus Caucasian. There was also the prediction that listeners whose performance on the implicit or explicit attitudes measures indicated that they were biased against African Americans or AAE would show this effect to a greater degree.

## Methods

### Overview

The study consisted of three parts: a speech rating task, a reaction time task designed to measure implicit racial bias, and a multipart questionnaire. In the speech rating task, participants listened to speech samples from children who speak African-American English (AAE), which were paired with pictures of children who were either African-American or Caucasian. The participants then rated the accuracy of the speech on a visual analog scale (VAS). The reaction time task was an Implicit Association Test (Greenwald et al., 1998) previously created for investigations of racial attitudes. The questionnaire was designed to elicit information pertaining to participants' knowledge of the features of AAE, their explicit attitudes about AAE and other issues related to dialects of English, and, for the trained listener group, information about their experience working with AAE-speaking clients. All participants completed all phases of the study, each of which is explained further below.

### Participants

A total of 40 adults participated in this study, divided evenly between the trained listener group and the untrained listener group. Participation was restricted to native monolingual English speakers over the age of 18 with no significant history of speech, language, or hearing disorders. Second language proficiency attained after childhood was not a disqualifying factor. Recruitment materials described the experiment as a "speech perception study" and included no mention of race or African-American English. Participants were compensated $10 for their time.

*Trained Listeners*

Twelve practicing pediatric Speech-Language Pathologists (11 female, one male) and eight graduate students (all female) in Speech-Language Pathology from the Minneapolis - St. Paul area formed the trained listener group. Seventeen out of 20 trained listeners self-identified as Caucasian, one as Asian-American, and two as "other." The racial, ethnic and gender characteristics of the trained listener group are approximately reflective of the regional demographics of the profession as a whole. The mean age of the trained listeners was 31.79 years (SD = 10.90), and the group reported a mean of 9.1 years of experience as Speech-Language Pathologists (SD = 11.32).

*Untrained Listeners*

Twenty individuals (10 female, 10 male) were recruited from the University of Minnesota community to form the untrained listener group. Data was excluded from one additional participant, who, after completing the study, reported that she had consistently reversed her answers during the speech rating task. Sixteen out of 20 untrained listeners self-identified as Caucasian, one as African-American, one as Asian-American, and two as "other." The mean age of the untrained listeners was 22.60 years (SD = 5.83). There was a significant age difference between groups, $t(37) = 3.31$, $p = 0.002$.

**Speech Rating Task**

*Stimuli*

The speech samples used in this study were collected at the University of Wisconsin - Madison as part of an interdisciplinary study of dialect mismatch supported by the Wisconsin Institutes for Discovery (Edwards, Gross, MacDonald, Brown, & Seidenberg, 2010). Speech samples were elicited from 109 children between the ages of four and

seven years, who were all speakers of African-American English. The particular samples used here were collected as part of the familiarization phase of a word comprehension experiment. The stimuli consisted of eight practice items and 18 pairs of pictureable words that would be familiar to the children: nine singular/plural pairs, such as "hat"/"hats," and nine monomorphemic word pairs differing in the presence of a final consonant cluster, such as "goal"/"gold." A female speaker of AAE recorded each word embedded in the carrier phrase "Say [word], please." These recordings were paired with color photographs and presented to the children on a touch screen laptop. The children wore a LENA recording device, which captured their productions. The recordings were later transcribed using Praat (Boersma & Weenik, 2005).

From the 44 monosyllabic words elicited from the speakers, words with four different final consonant sequences were selected: two singletons (final /t/ as in "cat;" final /l/ as in "bell") and two clusters (monomorphemic final /ld/ as in "cold;" bimorphemic final stop consonant plus plural morpheme - *s* as in "cats"). The final set of target items included 16 words, four with each of these four final consonant sequences, to include in the present study. These words are listed in Table 1 below. These words were excised from the larger recordings using Praat tools and saved as .wav files.

Ten individual productions of each of the sixteen target words were included in the speech rating task, yielding 160 total stimulus items. The stimuli were chosen such that production of the final consonant or cluster ranged from very accurate to very inaccurate across the instances of each word.

*Table 1: Target words used in the speech rating task.*

| Singleton Codas | | Cluster Codas | |
|---|---|---|---|
| */t/* | */l/* | */ld/* | *stop + plural -s* |
| bat | bell | bald | bats |
| cat | coal | build | books |
| coat | hole | cold | hands |
| hat | wheel | gold | hats |

So that listener's ratings of accuracy would reflect mainly the production of the final consonants, no tokens were chosen with perceptible errors or distortions affecting any part of the word other than the final consonant or consonants. Similarly, no selected recordings were of compromised quality due to muffled speech or background noise. Speech samples from 96 of the 109 talkers were included in the final study, with one or two words from selected each individual.

So that the listeners should impute different speaker races onto the various speech samples, each talker was paired with a photograph of either an African-American or Caucasian child of early elementary age. The pictures came from an online stock photo subscription site, as well as from a collection of images licensed to the University of Minnesota - Twin Cities (Eyewire Images, 2002). Since perceived speaker age has been shown (e.g., Drager, 2005; Munson, Edwards, Schellinger, Beckman, & Meyer, 2010; Schellinger, Edward, and Munson, 2010) to affect listener's judgments of speech accuracy, a pilot task was conducted where five respondents estimated the ages of the

children in the photographs. No significant difference in perceived age was found between the African-American and Caucasian-American children in the photographs.

The gender of the child paired with the speakers was determined by experimenter judgment. For the 96 talkers, 52 were paired with pictures of girls (26 African-American and 26 Caucasian) and 44 with pictures of boys (22 African-American and 22 Caucasian). For those speakers from whom two productions were selected, the same picture was paired with both samples of the individual's speech.

There were two versions of the experiment, such that every speech sample paired with an African-American child's face in the first version was paired with a Caucasian child's face in the second, and vice versa. The same set of images appeared in both versions. Within both the trained and untrained listener groups, half of the participants completed Version One and half completed Version Two, though the presentation order of individual test items was randomized for each participant.

*Procedure*

The speech rating task was programmed and executed using E-Prime experiment management software, version 1.2 (Schneider, Eschman, & Zuccolotto, 2002). This and the following portion of the study took place inside a sound-treated booth, where participants were seated in front of a 14-inch computer monitor placed approximately at eye level. The computer was equipped with a keyboard and mouse. The auditory stimuli were peak normalized and delivered to listeners over Realistic Nova 40 supra-aural headphones at a comfortable listening level of approximately 70 dB.

As some professional Speech-Language Pathologists in the trained listener group were unable to come to the lab to complete the study, these individuals were

accommodated by loading the experiment files onto a laptop computer. The experimenter visited these subjects in their homes or offices, where the study then took place. Seven individuals in the trained listener group elected to complete the study off site. For these participants, the procedure differed only in that the computer tasks were delivered on a laptop computer equipped with a 13-inch display, external mouse and Sennheiser HD 260 Pro headphones, and that the study took place in a quiet, private space at the participants' home or workplace rather than in the sound-treated booth.

Listeners read the instructions at their own pace on the computer monitor before beginning the task. The instructions informed the participants that they would be completing a study about the accuracy of children's speech production and that they would be seeing pictures of the children who produced the samples. On each trial, the text "Listen to the child say the word [WORD]" was displayed on the screen in 36-point Courier font. Next, the computer displayed the picture of the child previously paired with the speaker of the speech sample, while the sample was delivered over the headphones. Listeners were the instructed to rate the accuracy of the speech production on a visual analog scale (VAS) ranging from "Completely Accurate" on the left endpoint, to "Completely Inaccurate" on the right endpoint. The VAS is shown in Figure 1 below. For each trial, listeners indicated their perception of the child's speech accuracy by using the mouse to select the corresponding location along this continuum. Listeners were encouraged in the instructions to use the entire line in making their ratings, rather than simply selecting between the two endpoints.

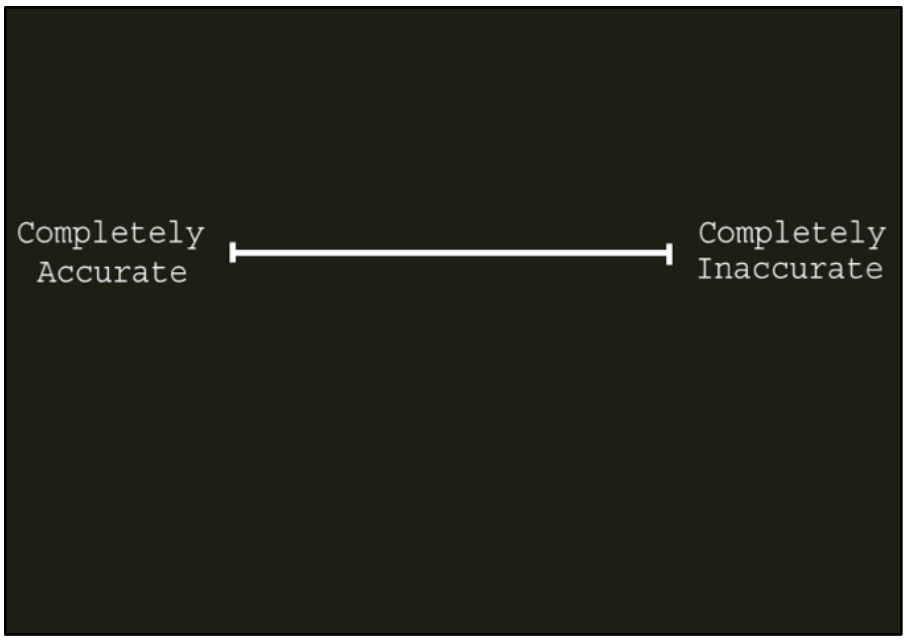


*Figure 1. The Visual Analog Scale used in the speech rating task. The horizontal line was 444 pixels long and centered 309 pixels from the left side of the screen.*

Participants completed four practice trials after reading the task directions and before beginning the task proper. The practice items consisted of words not targeted in the current study, which were paired with faces of Asian-American children. On every trial, the computer recorded the coordinates, in pixels, of the listeners' selections for later analysis.

**Implicit Association Task**

The current study used an Implicit Association Test (IAT) previously used by Babel (2009) to investigate the relationship between implicit racial biases, perceived speaker race, and speech perception. This particular IAT uses 20 stereotypically African-American names (associated with "BLACK," e.g. "Tyrone"), 20 stereotypically Caucasian-American names (associated with "WHITE," e.g. "Luke"), as well as two

other sets of 20 words, associated with "good" and "bad," respectively. The "good" words, such as "vacation," carry generally positive associations, while the "bad" words, like "vomit," carry negative associations. Babel's complete set of stimulus items, drawn from Greenwald et al. (1998), Dasgupta and Greenwald (2001), and Jelenec and Steffens (2002), can be found in Appendix A. Since Dasgupta and colleagues (2000) previous established that familiarity with the names used in the IAT did not affect performance, this factor was not controlled for in this study.

*Procedure*

Like the speech rating task, the implicit association task was programmed and executed using E-Prime experiment management software, version 1.2 (Schneider, Eschman, & Zuccolotto, 2002). Participants completed this task immediately following the speech rating task, in the same environment and using the same computer equipment as in the previous task.

The experimenter delivered instructions to the participants orally following the completion of the speech rating task. The "1" and "3" numeral keys on the keyboard were assigned to correspond to the category choices displayed on the left-hand and right-hand sides of the computer monitor, respectively. The experimenter instructed each participant to use only the index finger of his or her dominant hand to select between these two keys.

The task comprised five blocks, as illustrated in Table 2 below. To prevent fatigue, participants were allowed to pause between each task block.

*Table 2. Block design of the Implicit Attitudes Test. Test blocks are indicated in boldface.*

| Block | Name | Items | Concepts/Attributes |
|---|---|---|---|
| 1 | Target-Concept Discrimination | 20 | BLACK ---- WHITE |
| 2 | Associated Attribute Discrimination | 20 | good ----bad |
| **3** | **Combined Test** | **80** | **BLACK/good ---- WHITE/bad** |
| 4 | Associated Attribute Discrimination - Reversed | 20 | WHITE ---- BLACK |
| **5** | **Combined Test - Reversed** | **80** | **WHITE/good ----BLACK/bad** |

During the first, second, and fourth blocks, the two concepts (BLACK and WHITE, blocks 1 and 4) or attributes (good and bad, block 2) were displayed in the upper corners of the monitor, one per side. Randomly selected names (blocks 1 and 4) or words (block 2) were then presented in the center of the screen, and participants were instructed to press the button indicating the associated concept or attribute as quickly and accurately as possible. After each response, the word "Correct" or "Incorrect" appeared in the center of the screen. During the test blocks (blocks 3 and 5), the concepts and the attributes were both displayed in the upper corners of the screen, one above the other. Randomly selected names (to be categorized by concept, ignoring the attributes) and words (to be categorized by attribute, ignoring the concepts) appeared in the center of the screen.

Figure 2 below shows an example of the display during one of these test blocks. The two test blocks differed in which concept was paired visually with which attribute.

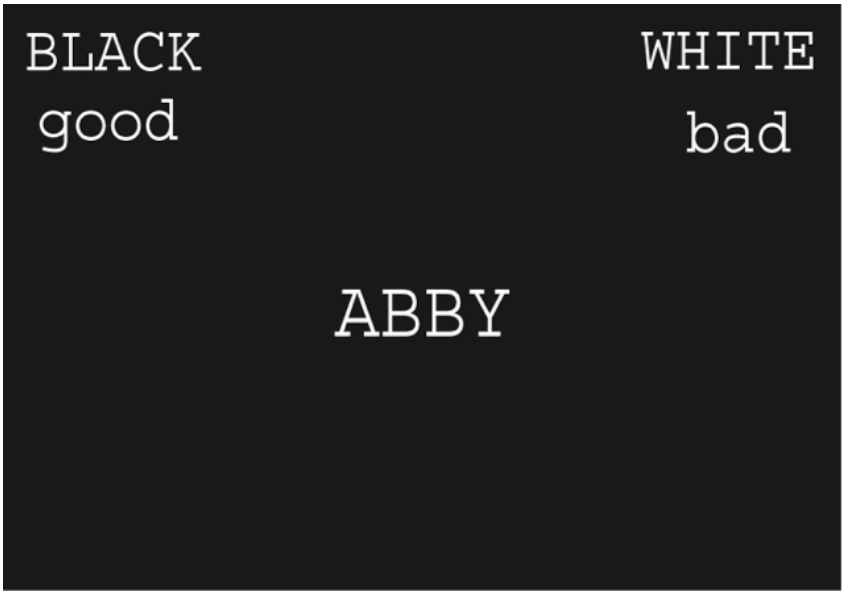


*Figure 2. Example item from one test block (third or fifth block) of the IAT.*

For each item, the computer recorded both the answer choice and response latency, which was later analyzed to determine each individual's implicit association score, related to the difference in response latency between the two test blocks (see Results section below).

**Participant Questionnaire**

A questionnaire was designed to elicit measures of participants' explicit knowledge of the features of AAE, their explicit attitudes relating to AAE, and, for the trained listener group, their experience working with AAE-speaking clients. The AAE Knowledge Survey and the Language Attitude Survey were programmed using E-Prime, so participants completed these parts of the tasks immediately following completion of

the speech rating task. They selected their responses by pressing the letter (AAE Knowledge Survey) or number (Language Attitude Survey) corresponding to each answer choice on the keyboard. The remainder of the questionnaire was completed in paper-and-pencil format following the completion of the implicit association task. This portion included demographic information, as well as questions for the trained listeners about their years of experience and the estimated percentage of their caseloads made up of speakers of African American English.

*AAE Knowledge Quiz*

To assess the participants' explicit knowledge of the phonological, syntactic, and morphological features of AAE, all listeners completed a 22 item, multiple-choice quiz. In some quiz items, participants chose the pair of words, from a field of three pairs, which would sound alike when produced by a speaker of AAE or Southern English. In other items, participants chose the phrase, wording, or expression most typical of AAE or Southern English, or interpreted the meaning of AAE expressions. The questions on this quiz originally appeared in Ford, Williams, Hicks, Hoover, Politzer, and McNair (1975), and can be found in Appendix B.

*Dialect Attitudes Survey*

As a complement to the measure of unconscious attitudes made using the Implicit Attitudes Test, there was also a measure of participants' explicitly held attitudes about AAE and speakers of non-mainstream English dialects. This took the form of a 25-item survey consisting of statements such as "African American English is lazy English" and "AAE would be inadequate for teaching subjects such as social studies or math." Respondents indicated their level of agreement with each statement according to a seven-

point Likert-type scale. The survey items were drawn from language attitude measures used by Vafadar and Utt (1993) and Blake and Cutler (2003), which appeared originally in Hoover (1996). A copy of the full survey is included in Appendix C.

**Debriefing**

After participants had completed all portions of the experiment, the experimenter debriefed them on the purpose of the study; specifically, it was explained that the study was examining issues related to race. This point had been intentionally withheld until the debriefing. The participants were then offered the chance to voluntarily withdraw their data from the study without changing their compensation. No participants chose to withdraw. All participants were then given a printed copy of the debriefing for their records, including information on how to withdraw from the study later. This text is included in Appendix D.

## Results

The following sections present the data gathered from the various measures as well as comparisons between the data obtained from the trained and untrained listener groups. The relationships between the results of the speech rating task and the listeners' knowledge of and attitudes toward African-American English are analyzed.

### Knowledge Measures

Participants' answers on the 22-item, multiple choice AAE knowledge quiz were recorded by the computer and scored by assigning one point to every correct answer, zero points to every incorrect answer, and calculating the percentage correct for every participant. The mean score, as well as mean scores for the two listener groups, are given in Table 3 below. There was a significant difference between groups; the trained listeners (mean = 0.784, SD = 0.093) as a group scored about 6% higher on the knowledge quiz than the untrained group (mean = 0.726, SD = 0.078; $t(38) = 2.154$, $p = 0.038$).

For the purposes of later analysis, the binary variable *High vs. Low Knowledge* was created, with participants whose scores fell above the median score of 77% classified as "high knowledge" and the rest as "low knowledge." The breakdown of these designations is also given in Table 3.

*Table 3. Mean scores on the 22-item AAE knowledge quiz for all participants and listener groups, and number of participants designated as "high knowledge" (score above median of 77%) and "low knowledge" (score at or below median).*

| Group | Mean Score | "High Knowledge" Listeners | "Low Knowledge" Listeners |
|---|---|---|---|
| All Participants | 75.1% | 15 | 25 |
| *Trained Listeners* | 78.4% | 11 | 9 |
| *Untrained Listeners* | 72.6% | 4 | 14 |

**Attitude Measures**

*Implicit Association Task*

The response latencies for each response made during the five blocks of the IAT test were recorded by the computer and analyzed by the experimenter to determine each individual's IAT score. Following Christy and Munson (2012) and Babel (2009), this study used the revised IAT scoring procedures outlined in Greenwald, Nasek and Banaji (2003). The resulting score, *d*, compares the difference in response latencies between the Target Attribute block (where "black" was paired visually with "good") and the Reversed Target Attribute block (where "white" was paired with "good"). A positive score indicates a pro-white bias, a negative score indicates a pro-black bias, and a score of zero is neutral.

The mean *d* for all participants, as well as mean scores for the two listener groups, are given in Table 4 below. There was not a significant difference between groups (Mean-Experienced = 0.324, SD = 0.378; Mean-Inexperienced = 0.216, SD = 0.389; $t(38) = 0.886$, $p = 0.381$). For the purposes of later analysis, the binary variable *High vs. Low*

*IAT Score* was created, with participants whose scores fell at or above the median *d* of 0.290 classified as "high IAT" and the rest as "low IAT." The breakdown of these designations is also given in Table 4.

*Table 4. Mean IAT score (*d*) for all participants and listener groups, and number of participants designated as "high IAT" (score above median of .290) and "low IAT" (score below median).*

| Group | Mean Score (Range) | "High IAT" Listeners | "Low IAT" Listeners |
|---|---|---|---|
| All Participants | 0.270 (-0.650 – 1.000) | 20 | 20 |
| *Trained Listeners* | 0.324 (-0.530 – 1.000) | 11 | 9 |
| *Untrained Listeners* | 0.216 (-0.650 - 0.970) | 9 | 11 |

*Dialect Attitudes Survey*

Responses on the dialect attitudes survey administered to measure participants' explicit attitudes toward AAE are illustrated on Figure 3 below, and the survey items are listed in Table 5. A summary score (the "explicit attitude test" or EAT score) for the dialect attitude survey was calculated by summing responses on all items, reversing the scoring on some items so that in all cases, a higher-numbered response indicated a position that was anti-AAE, pro-MAE, more prescriptive, or more likely to interpret dialect features as a disorder. Correspondingly, low-numbered responses indicated positions that were pro-AAE, more descriptive in nature, and more likely to interpret dialect features as a language difference.

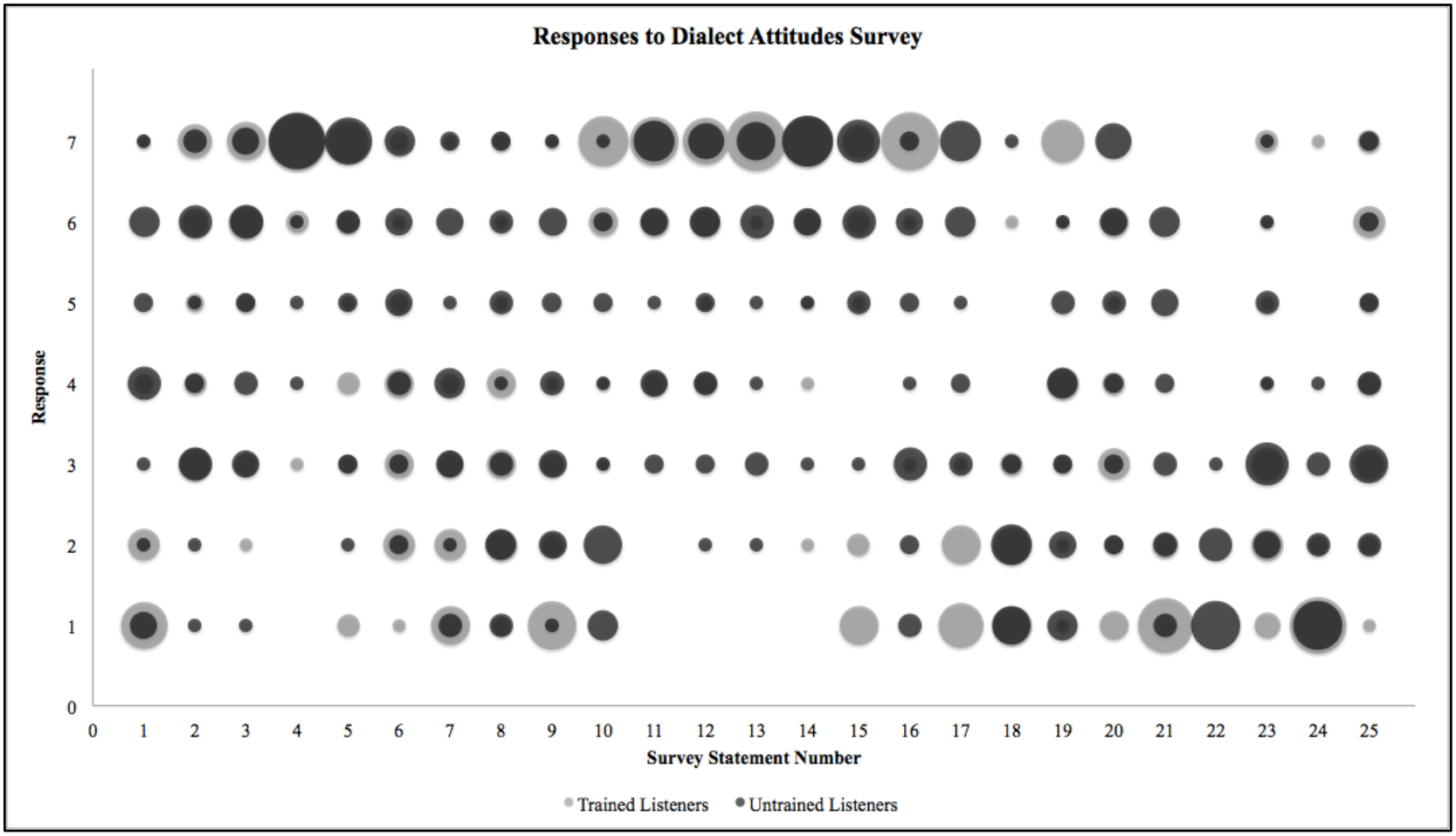


*Figure 3. Responses to the dialect attitudes survey. The area of the bubbles represents the proportion of 1-7 responses for each item. A response of 1 indicates complete agreement, while 7 indicates complete disagreement. Statements are listed in Table 5.*

*Table 5. Statements on the Dialect Attitudes Survey.*

| Item | Statement |
|---|---|
| 1 | African American English is subject to its own set of rules |
| 2 | African American kids would advance further in school if they didn't use African American English |
| 3 | African American students have language problems similar to those of students learning English as a second language |
| 4 | African American students should be taught in special classrooms with English as a Second Language (ESL) students |
| 5 | African-American English is lazy English |
| 6 | African-American English would be inadequate for teaching subjects such as social studies or math |
| 7 | Bidialectal education is the right of every child who does not speak the dominant language or dialect |
| 8 | Bilingual education is the right of every child who does not speak the dominant language |
| 9 | Federal funds should be used to support bi-dialectal education |
| 10 | Federal funds should be used to support bilingual education |
| 11 | I believe people who speak nonstandard dialects should receive speech and language therapy |
| 12 | I believe speakers of social dialects, whether disordered or not, should seek the services of a speech-language-pathologist in order to reduce the dialect |
| 13 | I believe that nonstandard English dialects occur because people don't completely learn the rules of Standard English |
| 14 | I believe that people who speak nonstandard dialects do so because they have language disorders |
| 15 | I believe that people who speak outside the norm of a particular community are displaying speech or language disorders |
| 16 | I believe that Standard English is one dialect |
| 17 | I believe that, from a practical standpoint, a non-standard dialect or language difference may be treated as if it were a disorder |
| 18 | In every language, people from different ethnic, social, and economic backgrounds speak differently |
| 19 | It is educationally sound to use a student's first language as a way of teaching that student the mainstream or dominant language of a community |

| Item | Statement |
|---|---|
| 20 | Mainstream English is dominant in schools and business because it is the best form of English |
| 21 | One purpose of school is to make certain that all students graduate proficient in Mainstream English |
| 22 | People speak differently in different situations |
| 23 | Some children do poorly in school because they do not speak Mainstream English |
| 24 | There are settings outside the classroom where African American English is appropriate |
| 25 | Using African American English as a tool to teach subjects to African American students would hurt their chances of learning Mainstream English |

The mean EAT score for all participants, as well as mean scores for the two listener groups, are given in Table 6 below. There was a significant difference between groups; the trained listeners (mean = 141.5, SD = 15.05) had higher EAT scores (indicating a more anti-AAE stance) overall than did the untrained listeners (mean = 127.9, SD = 17.86; $t(38) = 2.604$, $p = 0.013$). For the purposes of later analysis, the binary variable *High vs. Low EAT Score* was created, with participants whose scores fell above the median score of 136 classified as "high EAT" and the rest as "low EAT." The breakdown of these designations is also given in Table 6.

*Table 6. Mean EAT score for all participants and listener groups, and number of participants designated as "high EAT" (score above median of 136) and "low EAT" (score below median).*

| Group | Mean Score | "High EAT" Listeners | "Low EAT" Listeners |
|---|---|---|---|
| All Participants | 134.7 | 19 | 21 |
| *Trained Listeners* | 141.5 | 13 | 7 |
| *Untrained Listeners* | 127.9 | 6 | 14 |

## Speech Rating Task

Before beginning analyses, the x and y coordinates of all mouse clicks on the speech rating task were examined for outliers. One response from one listener was located at an extreme y value, indicating inattention to task, and was omitted from further analysis. All other data were included. When participants clicked beyond the endpoints of the line, located 87 and 530 pixels from the left side of the screen, those values were rounded up or down to 87 or 530 respectively. All x coordinates were then centered such that a click on the midpoint of the inaccurate-accurate continuum had an x value of 0, negative values indicated a rating to the left of the midpoint (more inaccurate), and positive values indicated a rating to the right of the midpoint (more accurate). These values were the dependent measures for all analyses in the following section.

## Analysis

### *Group Differences - All Target Words*

For the first set of analyses, listeners' accuracy ratings were averaged across all target words paired with pictures of African-American and Caucasian children, respectively. These data were submitted to three repeated-measure analyses of variance (ANOVAs), with mean accuracy rating as the continuous dependent measure, speaker race (African-American vs. Caucasian) as a within-subjects factor, listener group (experienced vs. inexperienced listeners) as a between-subjects factor, and one of the following additional between-subjects factors: (1) high vs. low EAT score, (2) high vs. low IAT score, and (3) high vs. low knowledge score. Each analysis, therefore, examined the individual and combined effects of listener group, speaker race, and one potential "bias factor" upon the listeners' judgments of the accuracy of the speech samples they heard. Table 7 below

displays the mean accuracy ratings within each cell for these analyses, and Table 8 summarizes the results of the ANOVAs.

*Table 7. Mean accuracy ratings (standard error) in the speech rating task.*

| | Experienced Listeners | | Inexperienced Listeners | |
|---|---|---|---|---|
| | African-American Speaker | Caucasian Speaker | African-American Speaker | Caucasian Speaker |
| **AAE Feature Knowledge** | | | | |
| *High Knowledge* | -77.99 (16.22) | -59.65 (15.96) | -115.48 (26.9) | -103.09 (26.47) |
| *Low Knowledge* | -90.36 (17.94) | -71.50 (17.65) | -78.13 (13.45) | -66.54 (13.24) |
| **Implicit Racial Attitudes** | | | | |
| *High IAT (pro-white)* | -70.86 (16.19) | -51.29 (15.95) | -75.5 (17.90) | -67.81 (17.63) |
| *Low IAT (pro-black)* | -99.07 (17.90) | -81.72 (17.63) | -93.86 (16.19) | -78.78 (15.95) |
| **Explicit Dialect Attitudes** | | | | |
| *High EAT (anti-AAE)* | -94.15 (13.83) | -73.05 (14.29) | -128.59 (20.35) | -104.30 (21.03) |
| *Low EAT (pro-AAE)* | -63.89 (18.84) | -50.00 (19.47) | -67.18 (13.32) | -60.79 (13.77) |

*Table 8. Summary of ANOVAs examining effects of listener group, speaker race, and knowledge/attitude factors on listeners' accuracy judgments. Significant effects are in boldface.*

| Analysis | SS | df | *F* | *p* |
|---|---|---|---|---|
| *1. High vs. Low Knowledge* | | | | |
| listener group | 3947.238 | 1, 36 | 0.711 | 0.405 |
| **speaker race** | **3638.978** | **1, 36** | **24.969** | **0.000** |
| knowledge | 2398.236 | 1, 36 | 0.432 | 0.515 |
| group x race | 169.415 | 1, 36 | 1.162 | 0.288 |
| group x knowledge | 9356.261 | 1, 36 | 1.685 | 0.202 |
| race x knowledge | 0.073 | 1, 36 | 0.001 | 0.982 |
| group x race x knowledge | 1.708 | 1, 36 | 0.012 | 0.914 |
| *2. High vs. Low IAT Score* | | | | |
| listener group | 209.713 | 1, 36 | 0.038 | 0.847 |
| **speaker race** | **4410.755** | **1, 36** | **31.125** | **0.000** |
| IAT | 9576.931 | 1, 36 | 1.729 | 0.197 |
| group x race | 247.648 | 1, 36 | 1.748 | 0.195 |
| group x IAT | 1063.957 | 1, 36 | 0.192 | 0.664 |
| race x IAT | 32.868 | 1, 36 | 0.232 | 0.633 |
| group x race x IAT | 113.974 | 1, 36 | 0.804 | 0.376 |
| *3. High vs. Low EAT Score* | | | | |
| listener group | 6949.958 | 1, 36 | 1.386 | 0.247 |
| **speaker race** | **4708.166** | **1, 36** | **38.024** | **0.000** |
| **EAT** | **27341.68** | **1, 36** | **5.454** | **0.025** |
| group x race | 20.296 | 1, 36 | 0.164 | 0.688 |
| group x EAT | 2908.186 | 1, 36 | 0.580 | 0.451 |
| **race x EAT** | **687.928** | **1, 36** | **5.556** | **0.024** |
| group x race x EAT | 125.097 | 1, 36 | 1.010 | 0.322 |

There was a significant main effect of imputed speaker race in all analyses, such that listeners rated productions paired with the face of an African-American child as less

accurate than those paired with the face of a Caucasian child. This effect recurred throughout all subsequent analyses.

There was also a significant main effect of EAT score, such that listeners with high EAT scores (indicating anti-AAE attitudes), rated the speech their heard as less accurate overall. EAT score also interacted with imputed speaker race, such that listeners with higher EAT scores showed a difference between accuracy judgments between samples paired with African-American versus Caucasian children, with the African-American children receiving lower ratings, while there was only a negligible race effect for listeners with lower EAT scores. This interaction is depicted in Figure 4 below.

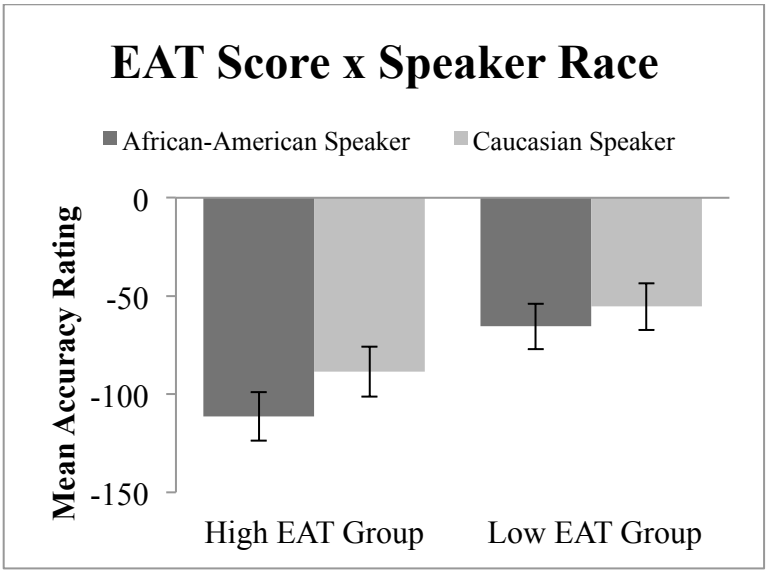


*Figure 4. Differences in mean listener accuracy rating by speaker race and high vs. low listener EAT score. Negative accuracy ratings (on the y-axis) indicate responses closer to "completely inaccurate" on the continuum, while positive ratings would indicate responses closer to "completely accurate." Error bars reflect the standard error of the mean.*

These analyses showed no effect of speaker group, IAT score, or knowledge of AAE features on listener judgments.

*Effects of Target Coda Type*

The data were submitted to an additional set of ANOVAs to examine differences in listeners' accuracy judgments across the different classes of target words included in the study (see Table 1). The first analysis compared words with singleton and cluster codas, the second compared words ending in /l/ and those ending in /t/, and the third compared targets ending in monomorphemic clusters (final *–ld*) with those ending in bimorphemic clusters (stop + plural *–s*). In all analyses, listener group (experienced vs. inexperienced listeners) was a between-subjects factor, and imputed speaker race (African-American vs. Caucasian) was a within-subjects factor. The results of these ANOVAs are summarized in Table 9 below.

*Table 9. Summary of ANOVAs examining effects of listener group, speaker race, and target coda types on listeners' accuracy judgments. Significant effects are in boldface. Effects approaching significance are italicized.*

| Analysis | SS | df | *F* | *p* |
|---|---|---|---|---|
| *1. Singleton vs. Cluster Coda* | | | | |
| listener group | 1737.52 | 1, 38 | 0.16 | 0.694 |
| **speaker race** | **9290.91** | **1, 38** | **33.02** | **0.000** |
| **coda** | **3545.12** | **1, 38** | **6.84** | **0.013** |
| group x race | 187.49 | 1, 38 | 0.67 | 0.419 |
| **group x coda** | **2248.95** | **1, 38** | **4.34** | **0.044** |
| race x coda | 256.34 | 1, 38 | 0.55 | 0.462 |
| group x race x coda | 57.74 | 1, 38 | 0.13 | 0.726 |

| Analysis | SS | df | *F* | *p* |
|---|---|---|---|---|
| *2. Final /l/ vs. Final /t/* | | | | |
| listener group | 4539.35 | 1, 38 | 0.40 | 0.533 |
| **speaker race** | **6872.92** | **1, 38** | **7.28** | **0.010** |
| coda | 338.99 | 1, 38 | 0.20 | 0.659 |
| group x race | 241.25 | 1, 38 | 0.26 | 0.616 |
| group x coda | 2626.91 | 1, 38 | 1.53 | 0.224 |
| *race x coda* | *1832.44* | *1, 38* | *2.88* | *0.098* |
| group x race x coda | 649.03 | 1, 38 | 1.02 | 0.319 |
| *3. Mono- vs. Bimorphemic Cluster* | | | | |
| listener group | 16.67 | 1, 38 | 0.00 | 0.970 |
| **speaker race** | **11962.26** | **1, 38** | **26.52** | **0.000** |
| **coda** | **271153.87** | **1, 38** | **100.06** | **0.000** |
| group x race | 793.79 | 1, 38 | 1.76 | 0.193 |
| group x coda | 3136.26 | 1, 38 | 1.16 | 0.289 |
| race x coda | 79.41 | 1, 38 | 0.14 | 0.715 |
| group x race x coda | 267.60 | 1, 38 | 0.46 | 0.503 |

As in the first set of analyses, there was a significant main effect of speaker race, such that listeners rated speech samples paired with pictures of African-American children as less accurate than those paired with pictures of Caucasian children.

There was a significant main effect of singleton vs. cluster coda type, such that listeners rated words ending in consonant clusters as less accurate than those ending in single consonants. This factor also interacted with listener group, as illustrated in Figure 5 below. The main effect of coda type is driven by the inexperienced listeners, who rated cluster targets as less accurate than singleton targets, while experienced listeners' ratings did not differ between coda types.

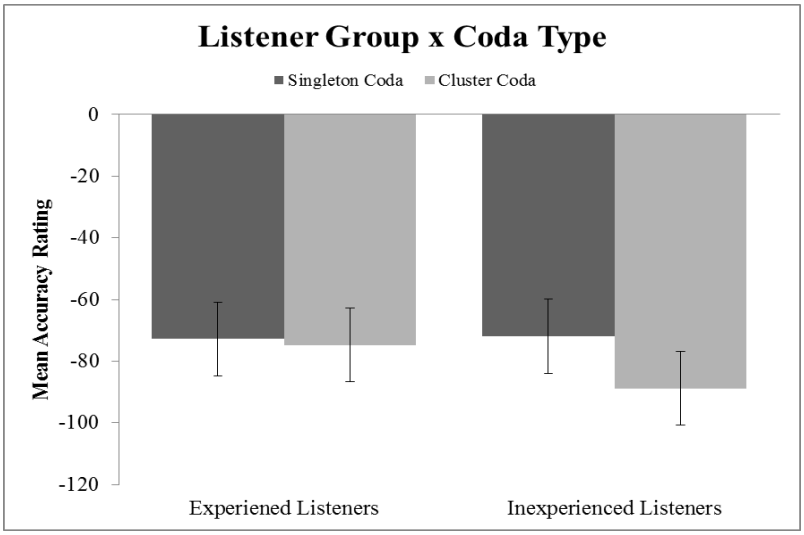


*Figure 5. Differences in mean listener accuracy rating by listener group and singleton vs. cluster coda type. Negative accuracy ratings (on the y-axis) indicate responses closer to "completely inaccurate" on the continuum, while positive ratings would indicate responses closer to "completely accurate." Error bars reflect the standard error of the mean.*

When comparing targets with final /l/, such as "bell," to those with final /t/, such as "cat," there was no main effect of coda consonant, however the interaction between coda consonant and speaker race approached significance. This interaction is plotted in Figure 6 below. There was a tendency for listeners to rate words with final /l/ as less accurate when paired with an African-American speaker, while for words ending in final /t/, the ratings were very similar for both African-American and Caucasian speakers.

Finally, in the comparison of words ending with monomorphemic consonant clusters (e.g., "cold") and bimorphemic clusters (e.g., "hats"), there was a significant main effect of cluster type, such that listeners perceived the bimorphemic targets as less accurate overall. This difference is illustrated in Figure 7 below.

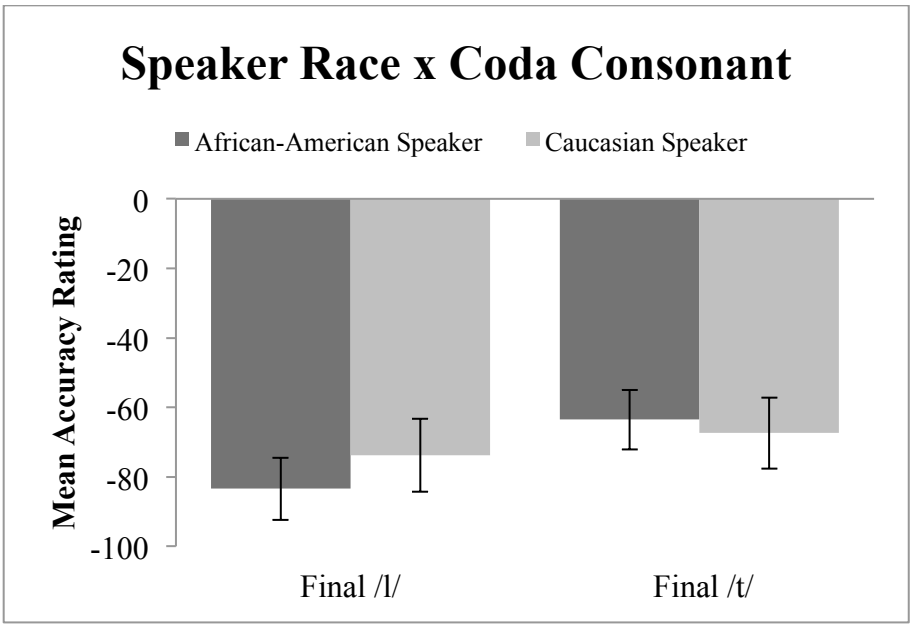


*Figure 6. Differences in mean listener accuracy rating by speaker race and final coda consonant (singletons only). Negative accuracy ratings (on the y-axis) indicate responses closer to "completely inaccurate" on the continuum, while positive ratings would indicate responses closer to "completely accurate." Error bars reflect the standard error of the mean.*

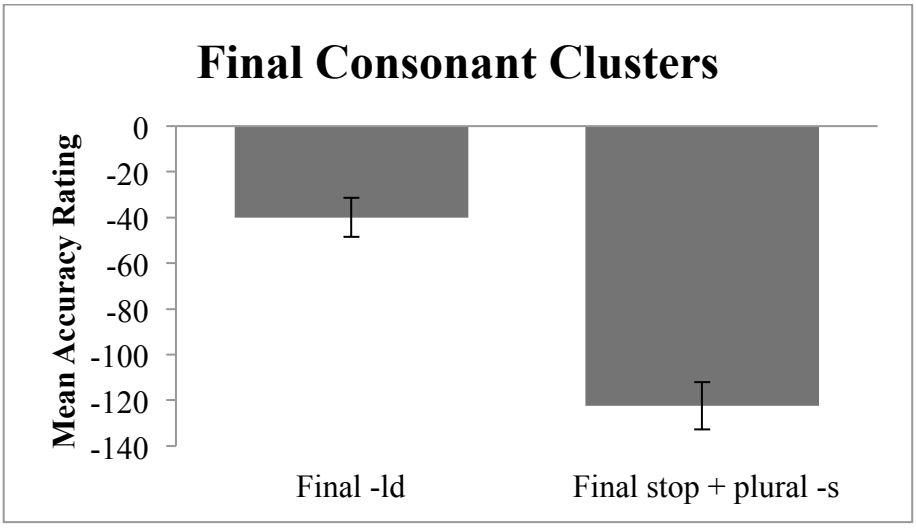


*Figure 7. Differences in mean listener accuracy rating by final consonant sequence (clusters only). Negative accuracy ratings (on the y-axis) indicate responses closer to "completely inaccurate" on the continuum, while positive ratings would indicate responses closer to "completely accurate." Error bars reflect the standard error of the mean.*

## Discussion

This study examined adult listeners' judgments of the accuracy of the speech of children who speak African American English (AAE). The aim was to investigate how listeners' knowledge of and attitudes toward African-Americans and AAE affected the perception of children's speech, and whether trained speech-language pathologists would judge the samples differently than inexperienced listeners. The main component of the study was an audiovisual perception experiment in which participants rated on a visual analog scale (VAS) the accuracy of words spoken by children they believed to be either African American or Caucasian, based on photographs that were paired with the speech samples.

In addition to the speech rating task, the listeners completed a quiz to assess their knowledge of the features of AAE, a survey on their explicit attitudes toward AAE and non-mainstream dialect use, and an Implicit Association Task designed to measure unconscious racial bias. It was found that the group of experienced listeners, which was made up of professional speech-language pathologists and advanced graduate students in that field, scored slightly higher on the AAE knowledge assessment than the inexperienced listeners. This is an intuitive result, given the professional training of an SLP. Interestingly, the experienced listeners also scored higher as a group on the measure of explicit dialect attitudes, which indicated more negative attitudes toward AAE and non-mainstream dialects of English. The item-by-item responses of both groups (see Figure 3) suggest that this difference may be due in part to differences in interpretation of phrases such as "speaking outside the norm of a particular community." Additionally, the

experienced listeners, most of whom are working professionals, provided more extreme responses to statements concerning "practical" measures.

It was predicted that the listeners would judge words to be less accurate overall when they were lead to believe the children who produced them were African American versus Caucasian. Indeed, such a race effect was apparent in all analyses, although it did not differ across listener groups as anticipated. There was a difference between listener groups when comparing ratings on words that ended with singleton codas (/l/ and /t/) with those ending in cluster codas (/ld/ and stop + plural *–s*). Inexperienced listeners rated the latter group of words as more accurate overall than the former, while experienced listeners judged both sets of words similarly. A possible interpretation of this result is that experienced listeners approached the task from an "articulation" standpoint, and rated words similarly regardless of whether morphology was affected, while inexperienced listeners regarded errors affecting the plural marker as generally more inaccurate; however, lacking more data this interpretation can only be conjecture.

Within just the group of stimuli ending in consonant clusters, those words ending in bimorphemic clusters (stop consonant + plural *–s*) were rated as less accurate overall than those words ending in monomorphemic (/ld/) clusters. This result must be interpreted cautiously, since objective accuracy of the stimuli was not controlled through *a priori* acoustic or perceptual criteria.

Within just the group of stimuli ending in singleton codas, there was a trend, approaching statistical significance, such that listeners judged words ending in /l/ as less accurate than those ending in /t/ when they were paired with African American faces, but not when they were paired with Caucasian faces. This is interesting, for while any

postvocalic consonant is eligible for deletion (or glottalization) in AAE, unreleased final /t/ is not uncommon in Mainstream American English as well, particularly in fast casual speech (Ladefoged & Johnson, 2011). This may render deleted final /l/ more "distinctively AAE-like" than glottalized or deleted final /t/. If this is the case, listeners may have been penalizing errors in final /l/ when paired with African American speakers to a greater degree because of the error's association with a non-standard dialect. While these stimuli were not constructed for the purposes of contrasting AAE and Mainstream American English phonological features, further research along those lines would help elucidate these findings.

It was predicted that listeners with less knowledge of AAE would judge words to be less accurate when they believed the speaker to be African American versus Caucasian. It was also predicted that listeners whose performance on the implicit or explicit attitudes measures indicated bias against African Americans or AAE would have a larger effect of speaker race on their ratings. The current results did not support any main effects of knowledge or implicit attitudes on listener judgments, or interactions between these factors and the perceived race of the speaker. There were, however, effects of explicit attitudes on accuracy judgments. Listeners whose answers on the explicit attitudes measure indicated more negative opinions towards AAE and non-mainstream dialects tended to rate the speech they heard as less accurate overall. On its own, this result could indicate that these participants tend to answer toward the extremes when using scaled instruments such as the VAS and the Likkert-type attitudes measure, but further analysis revealed that explicit attitudes also interacted with perceived speaker race. The interaction was such that while people with less positive attitudes towards AAE rated

words paired with African American speakers as less accurate than those paired with Caucasian speakers, listeners with more positive attitudes did not. This result is consistent with a theory that some listeners penalize African American talkers, consciously or unconsciously, for articulation "errors" that result from differences in AAE phonology. Such a tendency, if present, would be clearly at odds with ASHA's positions on what constitutes a speech difference versus a disorder. Further study is needed to shed more light on this possibility. Additionally, the issue of whether the observed differences are in fact conscious or unconscious is an intriguing one; an experiment eliciting listeners' thought processes and impressions of their own ratings would be an informative complement to the current results.

These findings are consistent with the basic conclusions of Christy and Munson: under certain conditions, adults' perceptions of children's speech can be influenced by the presumed race of the talker, and these differences are mediated by bias in the listener. The current work generalizes this finding to speech produced by AAE-speaking children and to speech with naturally occurring variation in accuracy. Although no effect of implicit attitudes was found in these data as Christy and Munson found in theirs, this may be due to differences in the available ranges of IAT scores and the way those scores were grouped for analysis. It must be noted that Christy and Munson found an interaction with listener IAT score only for /t/-final words that were modified to simulate a glottalized final stop; an analogous analysis was not possible with the current data.

Although the availability of speech samples from a large number of AAE-speaking children was an invaluable resource in this study, the scope of the research question was limited by the stimuli. As it was, all of the target words were eligible for consonant

deletion or cluster reduction in AAE, so listener judgments on developmental articulation errors versus AAE phonological features could not be readily contrasted. An experiment consisting of a balanced set of stimuli from both AAE and non-AAE talkers would be an excellent follow up to the current work. This study also leaves to future analysis the question of whether the effect of presumed speaker race shifts the entire distribution of listener ratings, or whether only certain parts of the distribution are affected. This discussion has operated under the former assumption, but more in-depth analysis is needed to examine this issue.

These results are relevant to the field of clinical speech-language pathology in that they enhance our nascent understanding of the many ways listener bias and known or assumed speaker characteristics affect speech perception and its clinical consequences. Ultimately, the hope for this line of research is to aid in the development of effective and efficient clinician training, where awareness, discussion, self-assessment and targeted interventions would serve to reduce bias and support the provision of appropriate and fair services for all children.

## References

American Speech-Language-Hearing Association. (2004). Knowledge and Skills Needed by Speech-Language Pathologists and Audiologists to Provide Culturally and Linguistically Appropriate Services [Knowledge and Skills]. Available from www.asha.org/policy.

Babel, M.E. (2009) *Phonetic and Social Selectivity in Speech Accommodation*. Dissertation. University of California, Berkeley.

Blake, R., & Cutler, C. (2003). AAE and Variation in Teachers' Attitudes: A Question of School Philosophy? *Linguistics and Education* 14(2): 163-194.

Boersma, P., & Weenik, D. (2005). Praat: Doing phonetics by computer (Version 4.3. 27). [Computer program]

Campbell-Kibler, K., (2006). *Listener Perceptions of Sociolinguistic Variables: The Case of –in(g)*. Dissertation. Stanford University.

Christy, A., & Munson, B. (2012) Presentation at the 13th Conference on Laboratory Phonology, Stuttgart, Germany.

Craig, H., Thompson, C., Washington, J., & Potter, S. (2003). Phonological features of child African American English. *Journal of Speech, Language, and Hearing Research*, 46, 623-635.

Dasgupta, N., & Greenwald, A.G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81(5), 800-814.

Dasgupta, N., McGhee, D.E., Greenwald, A.G., Banaji, M.R. (2000). Automatic Preference for White Americans: Eliminating the Familiarity Explanation. *Journal of Experimental Social Psychology* 36, 316-328.

Drager, K. (2005). From bad to bed: The relationship between perceived age and vowel perception in New Zealand English. *Te Reo*, 48, 55-68.

Edwards, J., Gross, M., MacDonald, M., Brown, M., & Seidenberg, M. (2010). Examing aspects of code-switching ability in children who speak African American English. Poster presented at SRCLD, Madison, WI.

Eyewire Images (2002). *Photography: Babies*. Getty Images, Inc.

Ford, J., Lewis, S., Hicks, S., Williams, D., Hoover, M. Politzer, R., & McNair, K. (1975). Tests of African American English for teachers of bidialectical students. In Jones, R. (1996). *Handbook of Tests and Measures for Black Populations*. Hampton, VA : Cobb & Henry Publishers, V.1, 367-381.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation - Second Edition.* American Guidance Service, Inc.

Greenwald, A.G., McGhee, D.E., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74:1464–1480.

Greenwald, A.G., Nosek, B., & Banaji, M.R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Algorithm. *Journal of Personality and Social Psychology*, 85:197-216.

Hay, J., & Drager, K. (2010) Stuffed toys and speech perception. *Linguistics* 48(4): 865-892.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review* 23(3): 351-379.

Hoover, M., McNair-Knox, F., Lewis, S., & Politzer, R. (1996). African American English attitude measures for teachers. In R. Jones, ed., Handbook of tests and measurements for Black populations, vol. I: 83-93. Hampton VA: Cobb & Henry.

Jelenec, P., & Steffens, M.C. (2002). Implicit Attitudes Toward Elderly Women and Men. *Current Research in Social Psychology*, 7(16), 275-292.

Ladefoged, P., & Johnson, K. (2011). *A Course in Phonetics – Sixth Edition*. Boston: Wadsworth, Cengage Learning.

Lindblom, B. (1990). Explaining variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), Speech production and speech modeling. Kluwer.

McGurk, H., & MacDonald J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Miller, J. (1981). Effects of speaking rate on segmental distinctions. In P. Eimas & J. Miller (Eds.), Perspectives on the Study of Speech. Hillsdale, NJ: Lawrence Erlbaum Associates.

Munson, B., Edwards, J., Schellinger, S., Beckman, M., & Meyer, M. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of *Vox Humana*. *Clinical Linguistics and Phonetics*, 24(4-5), 245-260.

Munson, B., & Seppanen, V. (2009). Perceived suggested gender affects ratings of the quality of children's spoken narratives. Presentation. [http://www.tc.umn.edu/~munso005].

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

Nygaard, L., & Lunders, E. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory and Cognition*, 30, 583-93.

Pendergast, K., Dickey, S., Selmar, J., & Soder, A. (1997). *Photo Articulation Test, Third Edition.* PRO-ED.

Perkall, J., & Klatt, D. (Eds.) (1986). *Invariance and Variability in Speech Processes.* Psychology Press.

Rickford, J., & Rickford, R. (2000). *Spoken Soul: The Story of Black English.*, New York: John Wiley & Sons.

Schellinger, S. K., Edwards J., Munson B., & Beckman M. E. (2008). Does "close" count in transcription as well as in horseshoes? Poster presented at the Child Phonology Conference.

Schneider, W., Eschman, A., Zuccolotto, A. (2002) E-Prime User's Guide, Version 1.2. [Computer Program] Pittsburgh: Psychology Software Tools.

Staum Casasanto, L. (2008). *Experimental Investigations of Sociolinguistic Knowledge.* Dissertation, Stanford University Department of Linguistics.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-4.

Thomas, E. (2007). Phonological and Phonetic Characteristics of African American Vernacular English. *Language and Linguistics Compass,* 1(5), 450-475.

United States Census Bureau. (2012). Most Children Younger Than Age 1 are Minorities, Census Bureau Reports. Press Release CB12-90.

United States Department of Education. (2001). 23rd Annual Report to Congress on the Implementation of the IDEA. Available from http://www2.ed.gov/about/reports/annual/osep/2001/index.html.

Vafadar, A., & Utt, H. (1992). A survey of Speech-Language-Pathologists' attitudes about, self-perceived knowledge of, and competency in dealing with social dialects: Language differences versus language disorders. *National Student Speech Language Hearing Association Journal*, 20, 65-72.

Walker, S., Bruce, V., O'Malley, C. (1995). Facial identity and facial speech processing: familiar faces and voices in the McGurk Effect. *Perception and Psychophysics*, 57(8): 1124-1133.

## Appendix A: Implicit Attitudes Test Stimuli

*Stimuli previously used by Babel (2009), and originally appeared in Greenwald et al. (1998), Dasgupta and Greenwald (2001), and Jelenec and Steffens (2002).*

| *Stereotypically African-American ("BLACK") Names* | *Stereotypically Caucasian-American ("WHITE") Names* | *Words Associated with "GOOD"* | *Words Associated with "BAD"* |
|---|---|---|---|
| Aaliyah | Abby | caress | abuse |
| Aijia | Amy | cheer | agony |
| Alonzo | Carson | diamond | awful |
| Andre | Claire | freedom | cancer |
| Dominique | Cody | friend | crash |
| Ebony | Colin | gentle | death |
| Jada | Connor | glorious | evil |
| Jamal | Dustin | happy | failure |
| Jazmine | Emily | health | filth |
| Latonya | Hannah | honest | horrible |
| Marquis | Heather | joy | hurt |
| Maurice | Jack | laughter | jail |
| Raven | Jake | love | murder |
| Shanice | Jenna | loyal | nasty |
| Temeka | Katherine | lucky | poverty |
| Terrance | Katie | paradise | rotten |
| Terrell | Logan | peace | sickness |
| Tiara | Luke | pleasure | terrible |
| Trevor | Madeline | rainbow | tragedy |
| Tyrone | Scott | wonderful | vomit |

## Appendix B: AAE Knowledge Quiz

*Questions from Ford, Williams, Hicks, Hoover, Politzer, and McNair (1975).*

1. Choose the missing word or words that an African-American English (AAE) speaker would be most likely to use to complete the phrase, "By the time I get back, you better _____ cleaned up this mess."

**(a) had**
(b) got to
(c) be done

2. To emphasize the fact that the action of the sentence, "Willie finished that work," was completed at a much earlier point in time, an AAE speaker would probably say:

**(a) Willie *been* finished that work.**
(b) Willie *did* finished that work.
(c) Willie *really* finished that work.

3. A close paraphrase of the AAE and Southern English phrase, "I 'mo go downtown" is:

I am anxious to go downtown.
**I am going to go downtown.**
I'm the one that's going downtown.

4. The AAE sentence "Don't nobody hit John," is best interpreted as meaning:

Nobody wanted to hit John.
Somebody hit John.
**Nobody hit John.**

***From the following sets, select the pair of words that may sound very much alike in the speech of AAE speakers and some Southerners.***

5.

| | |
|---|---|
| **(a) build** | **bill** |
| (b) boy | bop |
| (c) blimp | bloom |

6.

| | |
|---|---|
| (a) tin | twin |
| **(b) tag** | **tack** |
| (c) tot | tote |

7.

| | |
|---|---|
| (a) make | mall |
| **(b) messed** | **mess** |
| (c) mom | mop |

8.

| | |
|---|---|
| (a) Bob | cob |
| **(b) Bess** | **best** |

(c) ban bam

9.

**(a) roof** **Ruth**
(b) room rude
(c) row tow

10.

**(a) help** **hep**
(b) who hot
(c) hip hop

11.

(a) cow cot
**(b) Carl** **cart**
(c) Cal Carol

12.

(a) for fur
(b) five jive
**(c) film** **fill**

13.

**(a) bud** **but**
(b) reckon raccoon
(c) broom brim

14.

(a) toe tore
(b) time tie
(c) telegraph telegram

15.

(a) apple axle
(b) and ain't
**(c) asked** **axed**

Select the most pronounced AAE English, and sometimes Southern English, phrases in each of the following items.

17.

**(a) John a student**
(b) John dones student work
(c) John a study

18.

(a) readin' tests
(b) reading n' writing
(c) reading' tes'

19.

(a) He aimed kinda high
**(b) He be going to the store**
(c) Be you go?

20.

(**a) She seem tall**
(b) She be seem
(c) She talled

21.

(a) My mother, they
(b) My mother ised
**(c) My mother, she**

22.

(**a) three coat**
(b) Forthy dollars
(c) two-by-two

23.

(a) The money arrived
(b) Bob money
(c) Root monies

## Appendix C: Dialect Attitudes Survey

*Drawn from language attitude measures used by Vafadar and Utt (1993) and Blake and Cutler (2003), which appeared originally in Hoover (1996).*

Please respond to each of these statements by selecting number 1-7 on the keyboard, as follows:

1 – Entirely agree
2 – Mostly agree
3 – Slightly agree
4 – Neither agree nor disagree
5 – Slightly disagree
6 – Mostly disagree
7 – Entirely disagree

1. People speak differently in different situations.

2. In every language, people from different ethnic, social, and economic backgrounds speak differently.

3. Bilingual education is the right of every child who does not speak the dominant language.

4. Federal funds should be used to support bilingual education.

5. Some children do poorly in school because they do not speak Mainstream English.

6. Bidialectal education is the right of every child who does not speak the dominant language or dialect.

7. Federal funds should be used to support bi-dialectal education.

8. African American English is lazy English.

9. African American English is subject to its own set of rules.

10. African American kids would advance further in school without African American English.

11. Using AAE as a tool to teach subjects to African American students would hurt their chances of learning Mainstream English.

12. There are settings outside the classroom where African American English is appropriate.

13. It is educationally sound to use a student's first language as a way of teaching that student the mainstream or dominant language of a community.

14. AAE would be inadequate for teaching subjects such as social studies or math.

15. African American students have language problems similar to those of students learning English as a second language.

16. African American students should be taught in classrooms alongside English as a Second Language (ESL) students.

17. Mainstream English is dominant in schools and business because it is the best form of English.

18. One purpose of school is to make certain that all students graduate proficient in Mainstream English.

19. I believe clients with nonstandard social dialects should receive treatment.

20. I believe that speakers of nonstandard social dialects have language disorders.

21. I believe that nonstandard English dialects are the incomplete learning of Standard English rules.

22. I believe that, from a practical standpoint, a non-standard dialect or language difference may be treated as if it were a disorder.

23. I believe that Standard English is one dialect.

24. I believe that people who speak outside the norm of a particular community are displaying speech or language disorders.

25. I believe speakers of social dialects, whether disordered or not, should seek the services of a speech-language-pathologist in order to reduce the dialect.

## Appendix D: Participant Debriefing

Now that you have finished this experiment, we would like to tell you what we are studying. In this experiment, we are interested in whether people perceive speech differently depending on the race of the person who produced it. In the task in which you were rating children's speech, you sometimes viewed pictures of African-American children, and sometimes children who were Caucasian or Asian-American. We are interested in whether people rate speech differently in these different conditions. We are also interested in whether this tendency is related to people's experience perceiving speech, which is why we are comparing untrained people's perception to the perception of trained speech-language pathologists. We are also interested in how this interacts with people's unconscious perception of race, which is what we measured in the reaction-time task.

Our goal is not to label people or to judge their behavior. Instead, we are interested in finding the best way to assess children's speech production so that we can make sure that what people report is as close as possible to what the children do.

Now that we have told you what our purpose is, we would like to give you the opportunity to withdraw your data. You can do this without changing your compensation, and without affecting your relationship to the University or to the people who did this study. If you choose to leave your data in, please know that we will treat it with complete confidentiality. We will not link your name to your experimental results in any published report of this work. All of our analyses will be done using generic subject IDs. All of the records linking your name to the subject ID are kept in a locked cabinet in a locked research lab, and will be accessible only to the researchers who are working on this project. If you choose to leave your data in now, but change your mind later, you may still remove your data. Just contact Benjamin Munson at Munso005@umn.edu, or (612) 624-0304.