ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Race Identification in American English

Yolanda Feimster Holt,[a] Tessa Bent,[b] Melissa Baese-Berk,[c] and Kathrin Rothermich[a]

[a] Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC [b] Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington [c] Department of Linguistics, University of Oregon, Eugene

ARTICLE INFO

Article History:
Received April 4, 2024
Revision received June 24, 2024
Accepted August 14, 2024

Editor-in-Chief: Cara E. Stepp
Editor: Emily Beth Myers

https://doi.org/10.1044/2024_JSLHR-24-00228

ABSTRACT

Purpose: This study examined the race identification of Southern American English speakers from two geographically distant regions in North Carolina. The purpose of this work is to explore how talkers' self-identified race, talker dialect region, and acoustic speech variables contribute to listener categorization of talker races.

Method: Two groups of listeners heard a series of /h/–vowel–/d/ (/hVd/) words produced by Black and White talkers from East and West North Carolina, respectively.

Results: Both Southern (North Carolina) and Midland (Indiana) listeners accurately categorized the race of all speakers with greater-than-chance accuracy; however, Western North Carolina Black talkers were categorized with the lowest accuracy, just above chance.

Conclusions: The results suggest that similarities in the speech production patterns of West North Carolina Black and White talkers affect the racial categorization of Black, but not White talkers. The results are discussed with respect to the acoustic spectral features of the voices present in the sample population.

There is a large body of research on the grammatical and morphosyntactic differences between African American English (AAE) or African American Vernacular English, reported here as AAE, and White American English (WAE; see Lanehart, 2015, for a discussion on WAE in North America). In the last 20 years, research has compared the acoustic–phonetic properties of AAE and WAE spoken in the same geographic location (Childs & Mallinson, 2004; Eberhardt, 2009; Mallinson & Childs, 2004, 2007; Risdal & Kohn, 2014). This research is important because differences in word choice, word order, and verb use are easily identified linguistic variations that mark AAE as different from WAE. However, segmental, subsegmental, and suprasegmental differences such as vowel pronunciation, voice quality, and intonation (Thomas & Reaser, 2004) may occur below the level of listener awareness yet provide listeners with information they can use to classify talkers into racial groups. Therefore, even when AAE speakers use lexical and morphosyntactic features consistent with the WAE variety spoken in their community, listeners may be sensitive to a speaker's racial identity due to subtle acoustic–phonetic variations.

Correspondence to Yolanda Feimster Holt: holty@ecu.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Regional Variation in AAE

Although historically AAE was described as a singular dialect (see uniformity hypothesis Wolfram & Fasold, 1974) used primarily by Black Americans of historical African descent (Green, 2002; Lanehart, 2015; Wolfram, 2007), modern evaluations of AAE report regional speech variation (Thomas & Bailey, 2015; Thomas & Coggshall, 2007; Wolfram, 2007) in the North (Coggshall & Becker, 2009; Labov et al., 2016), in the West (Calder & King, 2022; King, 2016; King & Calder, 2020; Mengesha, 2022; Wassink, 2015), in Washington D.C. (Lee, 2011), and in the South (Bailey & Thomas, 2021; Cukor-Avila, 2001; Farrington, 2018; Fridland, 2003; Holt, 2018; Wolfram, 2019; Wolfram & Thomas, 2002). AAE vowel production can vary across regional dialects (Childs & Mallinson, 2004; Holt, 2018; Mallinson & Childs, 2004; Risdal & Kohn, 2014).

This regional AAE vowel variation results in similar but not fully aligned vowels between AAE and WAE speakers. One example of this pattern is back vowel fronting, a well-described vowel change observed in many communities across the United States (Labov, 2006; Thomas, 2002). Back vowel fronting is the production of the back vowels /u/ as produced in *whod* and /o/ as produced in *hoed* with a higher second formant (F2) value compared to the cardinal production of the vowel. The prevalence of back vowel fronting for WAE speakers in the United States is widespread (Labov, 2006). In contrast, there is limited evidence of back vowel fronting in Black talkers (Anderson, 2008; Holt, 2018; Thomas, 1989). When Black talkers participate in the fronting of /u/ and /o/, their participation varies by talker age and regional dialect (Anderson, 2008; Arnson & Farrington, 2017; Eberhardt, 2009; Thomas, 2002; Wolfram & Thomas, 2002).

The low back merger of the vowels in *cot* and *caught* is another widespread feature of WAE; however, AAE speakers are reported to be resistant to this change in vowel production. Low back mergers have been observed in geographically distant dialect regions, including New England (Clopper et al., 2005), Midland (Majors, 2005), the Western States (King, 2016; King & Calder, 2020; Labov, 2006; Wassink, 2015), and the Southern United States (Irons, 2007). Although the low back merger is observed throughout the United States, it is not a defining feature of Southern American English. Reported resistance has been observed with AAE speakers in Texas (Bernstein, 1993), Tennessee (Fridland & Bartlett, 2006), and numerous large centers across the United States, as surveyed in the Atlas of North American English (Labov, 2006). Again, an exception to this resistance was observed for AAE speakers in Pittsburgh (Eberhardt, 2009).

A third vowel with observed regional and racial distinctiveness is monophthongization of the diphthong /a͡ɪ/. In the Southern United States, both AAE and WAE speakers frequently produce this vowel as a monophthong (Childs & Mallinson, 2004; Thomas & Bailey, 2015), but the contexts under which this vowel is monophthongized are regionally and racially conditioned. For AAE talkers in the Western, Eastern Coastal, and Central Piedmont regions of North Carolina, /a͡ɪ/ is likely to be produced as a monophthong syllable (e.g., *tie*) before voiced stops (e.g., *tide*; Wolfram et al., 2000). However, only talkers in the Western mountains and Eastern Coastal Plain regions produce the monophthongal version preceding voiceless stops (e.g., *tight*).

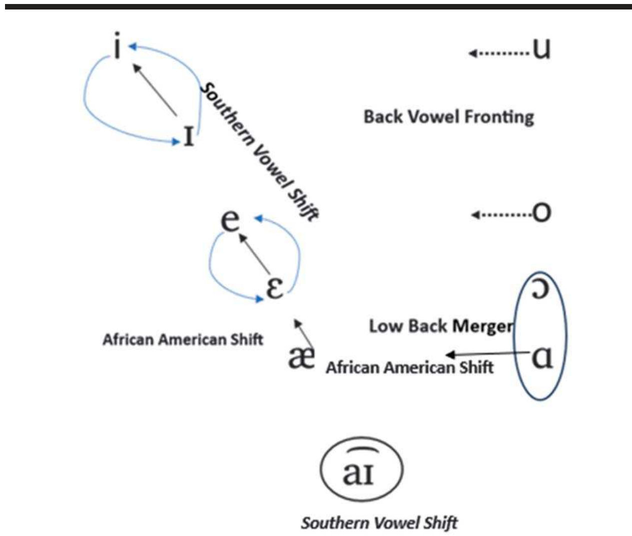
African American Shift

The African American shift (AAS) is a hypothesized vowel change occurring with the lax front vowels /æ/, /ɛ/, and /ɪ/ and the low back vowel /ɑ/. Thomas (2007) proposed that this vowel change phenomenon occurs solely in AAE speakers. In the AAS, the lax front vowels rise (increase in F2, decrease in first formant [F1]), and the low back vowel /ɑ/ lowers slightly and moves forward (increase in F1 and increase in F2). In the AAS, lax vowel changes occur with no concomitant change to tense vowel production, in contrast to the Southern vowel shift (SVS), where change is expected for both the tense and lax front vowels. In the SVS, the front tense and lax vowels move toward each other as the tense vowels become lax and lower (increase in F1, decrease in F2), whereas the lax vowels tend to move upward (decrease in F1, increase in F2). SVS may be initiated by the monophthongization of /a͡ɪ/. The monophthongization of /aɪ/ to /a:/ creates a gap in the vowel space. In the SVS, that gap is filled by the laxing and lowering of /e/ vowel and a concomitant tensing and raising of /ɛ/. The final step in the shift is tensing and raising of /ɪ/ and the laxing and lowering of /i/. When the SVS has moved to near completion, the tense lax pairs /i/−/ɪ/ and /e/−/ɛ/ either nearly or fully change places (Labov, 2006). A schematic of the vowel changes described, including SVS, AAS, low-back merger, and back vowel fronting, is presented in Figure 1, based on information from Labov et al. (2008). Although evidence of AAE participation in the AAS is limited, much of the data presented in support of the AAS are from Southern AAE speakers (Thomas, 2002, 2007; Thomas & Bailey, 2015; Thomas & Coggshall, 2007).

AAS and SVS

Both AAS and SVS are expected in the speech of Black talkers in North Carolina. Holt and colleagues (Holt, 2018; Holt et al., 2015; Holt & Rangarathnam,

Figure 1. Vowel change events occurring in North American English: the Southern vowel shift (curved lines), the African American shift (solid black lines), back vowel fronting (dashed lines), and the low back merger of open /o/ and the low back vowel. Based on information from Labov et al. (2008).



2018) investigated the distribution of vowel change events. They analyzed vowel production for AAE and WAE speakers from East and West North Carolina for participation in both the SVS and AAS. Holt (2018) showed participation in the SVS for both Black and White contemporary Western North Carolina talkers. A cross-generational comparison of Black and White talkers (Holt, 2011) found that the younger generation of West North Carolina White talkers decreased their participation in the SVS. In contrast, the younger generation of West North Carolina Black talkers maintained their participation in the SVS. In the East, neither Black nor White speakers participate in the SVS (Holt, 2018). The East North Carolina Black talkers had relatively raised front lax vowels with no concomitant change in tense vowel production, an indication of participation in the AAS (Holt, 2018). A subset of these data (Holt, 2011, 2018), along with data collected in Holt and Rangarathnam (2018), is used for the current perception experiments. Figures 2 and 3 show the descriptive representations of F1 and F2 at the 50% duration point for the front vowels *heed*, *hid*, *heyd*, and *head*; the back vowels *whod* and *hood*; andthelo w back vowel, as in *hod.*

Figure 2 (men) illustrates talker participation, or lack thereof, in the SVS relative raising and fronting of the tense vowels with concomitant backing and lowering of the lax vowels, the AAS relative raising of the lax vowels without lowering of the tense counterpart and lowering and fronting of the low back vowel as in *hod*, and the back vowel fronting of *whod* and *hoed*. West North Carolina men, Black and White, participate in the SVS, while East North Carolina Black men participate in the AAS. The West North Carolina Black males, indicated with black circles, show raising of the lax front vowels compared to their East North Carolina Black peers. West North Carolina Black male talkers produce /ɛ/ as in*head* much closer to the /e/ vowel as in *heyd*. The West North Carolina Black /ɪ/ vowel, as in*hid*, is fronter relative to /e/ and /ɛ/. The East North Carolina Black males have raised /æ/ as in *had* and /ɛ/ as in*head* relative to East North Carolina White peers although /ɪ/ vowel remains relatively low. All male talkers maintain relatively low production of /ɑ/ vowels as in *hod* with East North Carolina Black talkers, showing relatively greater lowering and fronting of the vowel consistent with participation in the AAS. The vowels /o/ as in *hoed* and /u/ as in *whod* remain fully back for the Black male talkers with relative fronting of the tokens by the White males.

Figure 2. Mean F1 and F2 of men for vowels involved in the Southern vowel shift, back vowel fronting, and African American shift (adapted from Holt, 2018). F1 = first formant; F2 = second formant; WAE = White American English; AAE = African American English.

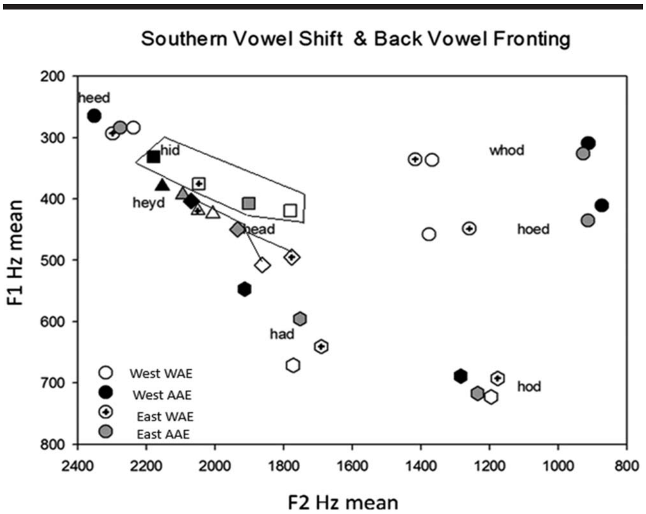


Figure 3. Mean F1 and F2 of women for vowels involved in the SVS, back vowel fronting, and African American shift. F1 = first formant; F2 = second formant; WAE = White American English; AAE = African American English.

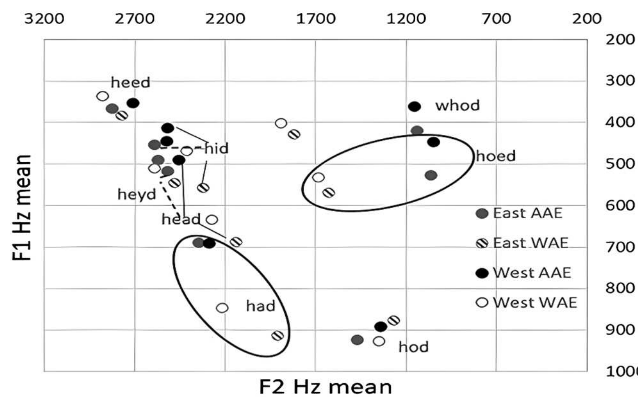


Figure 3 (women) illustrates a similar model of East North Carolina and West North Carolina AAE and WAE vowel production consistent with West talker participation in the SVS and East Black talker participation in the AAS. For West Black women, black circles show relative raising and fronting of /ɪ/ vowel asin*hid* and /ɛ/ as in *head*. Although close, /ɪ/ vowel is higher for West when compared to East Black women. Both Black talker groups raised /æ/ as in*had* compared to their White peers. The low back vowel (*hod*) remainslow forall talker groups, with East Black women showing greater lowering and fronting, consistent with participation in the AAS. Again, Black talkers maintain *whod* /u/ and *hoed* /o/ fully back compared to their White peers relative to the fronting of the vowels. The women's datapresented here include only a subset of the speakers used in the current experiment.

Voice Quality and Talker Race

Talkers who adhere to AAE speech production norms show some segmental and suprasegmental differences from speakers who use WAE. Differences in voice quality (e.g., fundamental frequency [*F*0], jitter, shimmer) between Black and White speakers have been previously reported (Andrianopoulos et al., 2001; Walton & Orlikoff, 1994). The influence of race on *F*0 shows no consistent significant difference between Black and White American speakers (Andrianopoulos et al., 2001; Walton & Orlikoff, 1994; Xue & Fucci, 2000), although one study found that lower *F*0 and greater frequency range for Black male talkers than White male talkers (Hudson & Holbrook, 1981). Some differences in prosodic characteristics have also been noted (Foreman, 1999; Holliday & Jaggers, 2015; McLarty, 2018; Spears, 1988).

Race Identification

Systematic differences across speakers from various socioethnic groups allow listeners to identify a speaker's race or ethnicity with above-chance accuracy from audio-only speech stimuli (Lass et al., 1978, 1979, 1980; Newman & Wu, 2011; Perrachione et al., 2010; Walton & Orlikoff, 1994). Some prior studies have focused on listeners' abilities to identify talkers as Black or White (Thomas & Reaser, 2004). Other investigations have included Hispanic, Asian, or Native American response categories (Newman & Wu, 2011; Purnell et al., 1999; Wong & Babel, 2017). For example, listeners accurately categorized the socioethnic identity of Black, White, Hispanic, and Asian-American speakers from New York City from a 60-word passage (Newman & Wu, 2011). Similarly, listeners from Vancouver accurately categorized Canadian speakers as White, Chinese, or East Indian based on semispontaneous sentence productions, although response accuracy was modulated by the match between the race and ethnicity of the speaker and the racial and ethnic composition of the listener's social network (Wong & Babel, 2017). Listeners can identify Black from White talkers at greater than chance accuracy in passages (Wolfram et al., 2000), sentences (Lass et al., 1978), and single vowels (Walton & Orlikoff, 1994). Race identification accuracy for speech that was synthetically manipulated, played backward, time compressed, low- or high-pass filtered, or monotonized remains robust (Thomas & Reaser, 2004). There is evidence the acoustic–phonetic characteristics of both AAE and WAE are influenced by regional variation; however, little research exists on how such regional variation or listener's familiarity with such regional variation may influence race identification accuracy.

Thomas and Reaser (2004) provided evidence that familiarity with a specific regional dialect influences listeners' racial categorization accuracy. In their study, listeners heard 20–30 s of connected speech from Black and White talkers from Hyde County (Outer Banks of North Carolina), an isolated enclave community with a unique regional dialect, as the test group. This speech and speech from a control group of Black and White speakers from the Piedmont and inland regions of North Carolina were presented to listeners (Thomas & Reaser, 2004). In both studies, the race identification results revealed that a talker's adherence to the unique (Hyde County) speech production norms strongly influenced the misidentification of Black talkers as White. Listeners from outside Hyde County were more likely to miscategorize the race of Black Hyde County talkers compared to White Hyde County talkers and Black and White talkers from central North Carolina. These groups were identified with greater than 80% accuracy. The extent to which speakers adhere to regional dialect norms appears to influence the accuracy of listener racial categorization. Perrachione et al. (2010) reported that Black talkers, producing phonetic variants associated with WAE including /u/ fronting, were miscategorized as White more frequently than Black talkers producing phonetic variants typically associated with AAE.

The specific acoustic–phonetic cues that listeners employ to make racial categorization judgments have not been fully determined, but a range of studies have identified potential cues (see Thomas & Reaser, 2004, for a review). Some research has pointed to differences in suprasegmental characteristics, including rhythm, intonation, and voice quality, which can used by listeners (Holliday & Jaggers, 2015; McLarty, 2018; Thomas & Lanehart, 2015) to identify speaker race. One method for identifying the importance of these features is to synthetically manipulate the stimuli (e.g., low-pass filtering, time compression, time reversal, monotonization) so that listeners are provided with access to specific acoustic–phonetic information. The results of these studies suggest that segmental and suprasegmental cues are important for race identification (Thomas & Reaser, 2004). Of relevance to the current study, listeners have been shown to attend to vowel cues in their race identification judgments (Bryden, 1968; Graff et al., 1986; Purnell et al., 1999; Walton & Orlikoff, 1994). For example, Purnell et al. (1999) showed that listeners could identify the race of a speaker from a single word, *hello*. Using a single speaker replicating the speech of a Black, White, and Hispanic/Latino speaker (code-switching), listeners categorized the speaker as the intended guise—Black, White, or Latino—with greater-than-chance accuracy. An acoustic analysis of the word *hello*, produced in each guise, revealed four acoustic cues that listeners may have used to make their dialect judgments: F2 in the lax vowel /ɛ/,pitch peak (*F*0), duration of the first syllable in *hello*, and harmonic-to-noise ratio (HNR). Of these four parameters, only F2 was found to statistically distinguish between the three dialects. A more recent study using these stimuli (Scharinger et al., 2011) suggested that the extraction of these acoustic differences and the resultant categorization of the guises happens very rapidly and pre-attentively. It must be noted that in this mixed guise experiment, a single male talker produced all three sets of tokens; therefore, the results may not be generalizable to experiments in which the stimuli for different socioethnic categories are produced by multiple speakers. Note that Purnell et al. compared AAE, a nonstandardized form, to Standard American English. Comparing nonstandard forms such as AAE to Standardized American English is an imbalanced comparison. Comparisons of nonstandard forms of speech used by Black and White talkers from the same geographic region or circumstance such as the works of Thomas and Reaser (2004), Walton and Orlikoff (1994), and Hawkins (1992) among many others may provide greater insight into the mechanisms listeners use to categorize a talker's race than comparing AAE and other nonstandard dialects to Standardized American English.

Current Work

In the current work, we continue the investigation of listeners' perception of race, with a focus on talker's regional variation and listener's location. Listeners heard 10 /h/–vowel–/d/ (/hVd/) words produced by age- and gender-matched groups of Black and White speakers from East and West North Carolina. The listeners were from either the Southern (from Eastern North Carolina) or the Midland (from central Indiana) dialect regions. The tokens presented were selected from previously analyzed data sets (Holt, 2011, 2018; Holt & Ellis, 2018). Previous acoustic spectral analyses (Holt, 2011, 2018; Holt & Rangarathnam, 2018) of the speech tokens presented to listeners in this experiment revealed that West Black talkers aligned with West White talkers in vowel production. Measurements included F1 by F2 in Hz for vowel location, vowel-inherent spectral change, and spectral rate of change. West Black men and women's alignment with White peers contrasted with that of East Black talkers who did not show a similar alignment of vowel production with East White peers. Based on previous work as described above, we posed the following research questions:

Q1. Does talker adherence to regional vowel productions that vary according to socioethnic/racial expectations affect listener race identification accuracy?
We hypothesized that listeners would be less accurate at race identification for West Black talkers whose vowel productions have already been shown to align more closely with the regional vowel productions of their WAE peers than for East Black and White speakers whose speech varies along socioethnic/racial lines.

Q2. Does the listener's dialect region—a proxy for dialect familiarity—influence race identification accuracy?
We hypothesized that Southern (North Carolina) listeners would show an advantage in race identification accuracy compared with Midland (Indiana) listeners.

Q3. Does gender influence race identification accuracy?
In general, West Black talkers showed greater alignment in vowel production with White peers than East Black talkers (Holt, 2018; Holt & Ellis, 2018). Therefore, we hypothesized that race identification accuracy may be influenced by interactions among talker sex/gender, race, and regional dialect.

Method

Listeners

There were two groups of listeners: current Indiana residents (n = 44; Midland listeners) and current East North Carolina residents (n = 28; Southern listeners). All listeners passed a pure-tone hearing screening at 25 dB for 250 Hz and 20 dB for octave intervals between 500 and 8000 Hz. The demographic characteristics of the participants are presented in Table 1. Most Southern listeners indicated their U.S. regional dialect as Southern (n = 27), with one participant indicating the same value for exposure to all regional dialects of American English. Midland listeners rated their exposure to the U.S. dialects, as described in Table 1.

Stimuli

The presented words were previously collected and analyzed for consistency with regional sound change (SVS) and socioethnic vowel change (AAS; Holt, 2011, 2018). Talkers self-identified as Black or White. As illustrated in Figures 2 and 3 and described in the paragraphs surrounding the figures, East and West White and Black talkers show consistency in vowel production by region and race. The East Black talkers are participating in the AAS but no aspects of the SVS. The West Black talkers are participating in some but not all aspects of the SVS. The West Black talkers are not participating in the AAS. The East White talkers show back vowel fronting but no front vowel changes consistent with the SVS. The West White talkers show back vowel fronting and changes to the front vowels consistent with the SVS. Listeners were presented with previously collected recordings of talkers producing the *hVd* words *heed*, *hid*, *head*, *had*, *hayed*, *hide*, *hood*, *howed*, *hoyed*, and *whod*. Each listener heard 24 productions of 10 /hVd/ words for a total of 240 different stimuli. Listeners heard two presentations of each word

Table 1. Listener demographic information and dialect exposure.

Listener group	Midland n = 44 (Indiana)	Southern n = 28 (North Carolina)
Age (in years)	average 21 (range: 18–30)	average 21 (range: 18–30)
Sex (self-identified)	18 males	13 males
	26 females	15 females
Ethnic group (self-reported)		
Hispanic/Latino	3	0
Not Hispanic or Latino	39	28
Prefer not to answer	2	0
Race (self-reported)		
White	31	15
Black or African American	7	12
Multiple races	3	1
Asian American	3	0
Other	1	0
U.S. regional dialect (self-reported)		
Current location	23	27
Northern	13	0
Some other regional dialect of American English	8	1
Dialect exposure for current location (1 = *no exposure* through 5 = *daily at home exposure*)	M = 4.5, range: 3–5	M = 4.6, range: 1–5

Note. For both Southern and Midland listeners, the predominant regional dialect exposure was consistent with the listener's physical location: Midland for Indiana and Southern for North Carolina.

from each speaker, for a total of 480 presentations. Talkers were evenly divided by sex, race, and geographic dialect region: 12 were from East North Carolina (Pitt County, representing the Coastal Plain dialect) and 12 from West North Carolina (Iredell County, representing the Piedmont dialect). The two dialect regions were geographically distant and separated by 370 km. The words presented to listeners were chosen to highlight regional and socioethnic differences in speech production between the East and West talkers and the Black and White talkers.

West talker speech is consistent with SVS, as identified by Labov et al. (2008). Holt (2018) previously evaluated the stimuli and established that Black West talkers, both men and women (Holt, 2011), participated in aspects of the SVS and produced vowels in a manner consistent with West White peers.

East speakers, neither Black nor White, participate in the SVS, and the vowel systems of Black and White speakers remain relatively distinct (Holt, 2018). Additionally, the East Black talkers, both men and women, show raising of the front lax vowels, consistent with participation in the AAS, as illustrated in Figures 2 and 3.

All speakers reported that they were typically developing with no history of speech or hearing disorders. At the time of recording, the speakers were free from colds and other upper respiratory diseases. The talkers were between 20 and 40 years of age (M =27) at the time of recording (Holt, 2011, 2018).

Procedure

Prior to the start of the experimental tasks, study procedures were approved by the East Carolina University Institutional Review Board (IRB) UMCIRB 18–000727. All listeners completed an IRB-approved consent form, protected health information form, demographic questionnaire, and hearing screening. The experimental assessment was conducted in a sound-attenuated booth (up to four listeners at a time) in Indiana and (one at a time) in North Carolina. All listeners completed two experimental tasks, a word identification task and a race identification task. Stimulus presentation was randomized. For each task, listeners were presented with two blocks of 240 trials. In each block, all 240 stimuli were presented to listeners. Therefore, listeners heard each speaker producing the target stimulus item 4 times (twice in the race identification task and twice in the word identification task). The focus of the current work is on listener ability to correctly identify talker race; therefore, only the protocol and results of the race identification task are reported in this article. The listeners were given breaks every 60 trials within each block. This break required listeners to pause for 20 s. Between blocks in each experimental task, listeners were required to take a 2-min break. The listeners were provided with a longer break between the two tasks.

At the start of each trial of the race identification task, a fixation cross was presented in the center of the screen for 500 ms, followed by a blank screen for 500 ms to orient the listener to the task and the location of the response boxes on each screen. After this orientation, two boxes appeared on the screen: one labeled Black and the other labeled White. One of the 240-word stimuli was played simultaneously with their appearance. Listeners were instructed to click the appropriate box to categorize their perceived race. Once the listener's selection had been made, the next trial began. Listeners did not have a time limit for their response entry, did not receive feedback on the accuracy of their responses, and could not replay the stimulus. For Midland listeners, each word was presented binaurally over Sennheiser HD280 Pro headphones at an SPL of approximately 68 dB. Southern listeners heard stimuli over Sennheiser HD280 Pro headphones at a listener-defined comfortable listening level of 45–70 dB SPL. Stimulus presentation and response recording were automatically controlled by a program written in PsychoPy (Peirce, 2007) on a Mac Mini 2.4 in Indiana, and on a Dell Precision 3200 in North Carolina. The total test time was 1.5 hr. Indiana participants were compensated for $15 for their time. Participants from North Carolina were volunteers, as no funds for payment were available at the time of the experiment.

Analysis

Listener response data were analyzed using mixed-effects logistic regression. Listener accuracy was defined as the correct identification of a speaker's race. Listener accuracy was used as the dependent variable. The models described below include talker race (i.e., Black vs. White), talker dialect (i.e., East vs. West), and the interaction between these two factors as fixed effects, and both factors were dummy coded. These factors capture the primary research questions. Additional models included talker gender and listener location to investigate the impact of these factors on race identification accuracy. Other possible interactions were found to be insignificant and are not included here. The presented models included random intercepts for speakers, listeners, and spoken words. The models using random slopes failed to converge. Therefore, the random slopes were not included in the model. Below, we describe the results of the models and present figures that illustrate these results.

To further evaluate the relationship between the acoustic characteristics of speakers' voices and race identification, principal component analysis (PCA) was completed. This procedure was chosen to identify acoustic characteristics that were significantly associated with race identification. As the acoustic spectral features collected for analysis are derived from the harmonic and inharmonic (noise) components of the vocal signal, strong relationships between the variables are expected. To account for these strong relationships, a regression analysis was performed to assist in identifying the relevant factors prior to submitting the variables to the PCA.

Results

Talker Race, Talker Dialect, and Their Interaction

Below, we present three figures, each of which further subdivides the data. Figure 4 shows the accuracy results as a function of the talker race and talker dialect. In this figure, listeners are more accurate at correctly identifying White talkers than Black talkers and are more accurate at identifying speakers from East North Carolina than West. This effect appears to be driven by Black speakers from the West North Carolina dialect. These were identified much less accurately than any other group of speakers. The statistics confirmed this observation.

Table 2 summarizes the full model. Both talker race and talker dialect significantly contributed to model fit ($z = 2.197$, $p = .028$; $z = -6.601$, $p < .0001$). Critically, the interaction between these two factors also significantly contributed to model fit ($z = 4.173$, $p < .001$). Taken together, these results suggest that both the talker dialect and talker race significantly improve the model fit.

Figure 5 shows the accuracy results for identification as a function of the talker dialect, talker race, and listener location. Southern listeners showed a slightly higher race identification accuracy than Indiana listeners. As expected, and according to the self-reports of dialect exposure, Midland listeners have fewer opportunities for exposure to North Carolina dialects than North Carolina listeners. The results of the mixed model provide additional support

Figure 4. Mean race identification accuracy (proportion correct) for Black talkers (left) and White talkers (right) and by dialect with East talkers in white and West talkers in gray.

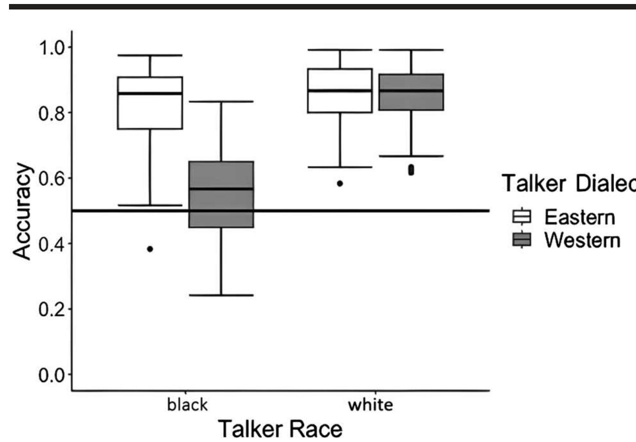


Table 2. Predictors.

Variable	Accuracy		
	Odds ratios	CI	*p*
(Intercept)	4.45	[3.24, 6.10]	**< .001**
Talker race [White]	1.53	[1.05, 2.22]	**.028**
Talker dialect [West]	0.28	[0.20, 0.41]	**< .001**
Listener location [North Carolina]	1.21	[1.01, 1.46]	**.044**
Talker gender [Woman]	1.28	[0.98, 1.68]	.065
Talker race [White] × Talker dialect [West]	3.11	[1.82, 5.29]	**< .001**

Note. Bold indicates statistically significant difference. CI = confidence interval.

for listener dialect exposure as a positive factor in talker race identification. Listener location significantly improved model fit (z = 2.017, p = .0437), suggesting that listeners with greater exposure to the target dialects more accurately identified a speaker's race than listeners with less exposure.

Figure 6 shows the listener accuracy as a function of talker race and talker dialect, and speaker gender. The figure illustrates the consistency with the overall pattern of identification accuracy for both men and women. Again, greater, but not statistically significant, variability is noted in race identification for West Black women than for West Black men. The results of the statistical analysis support this observation, as talker gender did not significantly improve the model fit (z =1.847, p = .0647).

Voice Quality Spectral Acoustics

The vowel portion of each word, *heed*, *hid*, *head*, *had*, *hayed*, *hide*, *hood*, *howed*, *hoyed*, and *whod*, was manually identified and marked in a Praat text grid and then submitted to the automatic voice analysis program VoiceSauce (Vicenik et al., 2024) using standard settings for parameter estimation as outlined in the program's manual. The following 26 variables were extracted pitch, $F0$, the first three formants (F1, F2, and F3), the first three harmonics' amplitudes calculated at the spectral peak maximum as estimated by $F0$ (H1*, H2*, and H3*), the first three formants amplitudes (A1*, A2*, and A3*) measured as the mean corrected value of formants and bandwidths every 20 ms, the relative corrected amplitudes of the first and second harmonics, (H1*–H2*), the relative corrected amplitudes of the second and fourth harmonics (H2*–H4*), corrected harmonic amplitudes using formant frequencies and bandwidths (H1*–A1*, H1*–A2*, H1*–A3*), and the spectral slopes from the third harmonic to the harmonic nearest 2 kHz in frequency (H3*–2k*), the spectral slopes from the fourth harmonic to the harmonic nearest 2 kHz in frequency (H4*–2k*), the root-mean-square of energy (Energy) calculated in the variable window equal for each five pitch pulses, cepstral peak

Figure 5. Race identification accuracy (proportion correct) is on the *y*-axis. Talker dialect is East in white and West in gray. Talker race is indicated on the bottom *x*-axis with Black on the left and White on the right in both panels. Listener location is indicated on the top *x*-axis with Midland listeners on the left and Southern listeners on the right.

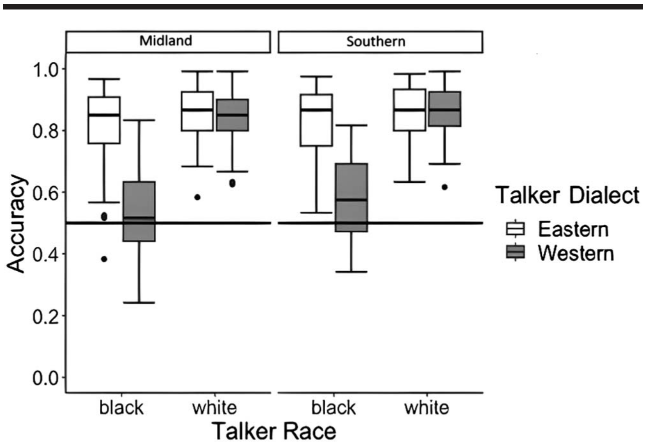
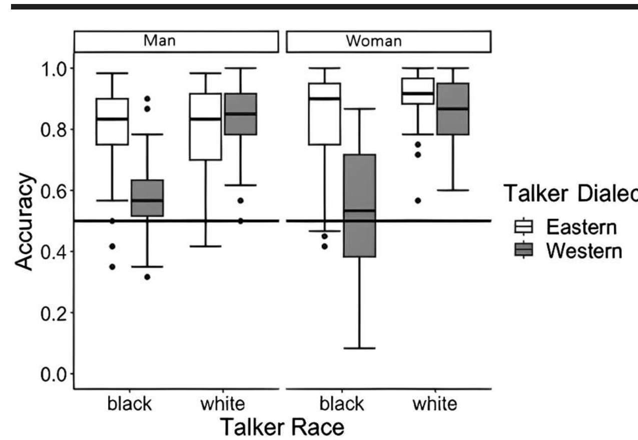


Figure 6. Race identification accuracy (proportion correct) by talker gender (men on the left and women on the right). Within each gender panel, Black talkers are on the left and White talkers are on the right. Talker dialect is shown by color: East in white and West in gray.



prominence (CPP), HNR from 500 to 3500 Hz (HNR 05 below 500 Hz, HNR 15 below 1500 Hz, HNR 25 below 2500 Hz, HNR 35 below 3500 Hz) subharmonic ratio (SHR), and strength of excitation (SoE). Vowel duration (Dur) in milliseconds was also calculated from the onset of the first positive pitch peak to its offset. Using the method described by Lee (2018), the acoustic variables were classified categorically, as illustrated in Table 3.

Acoustic variables were inspected for outliers, defined as values of more than 3 *SD*s from the arithmetic mean. Outliers were not identified in the data. Next, zero values were removed, as they represent variables that could not be measured owing to pitch tracking errors or an inaccurate measure of *F*0. All values were normalized between 0 and 1 prior to analysis.

We first added voice quality measures to the logistic mixed models in the manner described above; however, the results were not interpretable, largely because none of the individual predictor variables were significant alone. This is likely because race categorization is not reliant on a single acoustic factor but a constellation of factors. Therefore, we submitted the values to forward and backward stepwise regression analysis in SPSS (Version 26.0) to identify significant variables in the equation. Regression equations were completed using race identification as a predictor variable. The backward regression model revealed Dur, H1*, H2*, A3*, H2*–H4*, H4*–2k*, CPP, Energy, HNR 05, HNR 15, HNR 25, HNR 35, SHR, *F*0, F1, F2, F3, and SoE, whereas the forward regression model revealed Dur, A3*, H1*–H2*, H2*–H4k*, H4*–2k*, CPP, Energy, HNR 05, HNR 15, HNR 25, HNR 35, SHR, F1, F3, and SoE. Nineteen of the 26 variables derived from Voice Sauce contributed significantly to the model. The variables Dur, H1*, H2*, A3*, *F*0, F1, F2, F3, SoE, Energy, SHR, CPP, HNR 05, HNR 15, HNR 25, HNR 35, H1*–H2*, H2*–H4*, and H4*–2k* were selected for PCA.

Next, the Kaiser–Meyer–Olkin (KMO) Test for Sampling Adequacy (IBM, 2021) was run to determine if our data were suitable for PCA. The KMO value was 0.58, indicating marginal suitability for analysis. To test the suitability of the data for a factor analysis, the Bartlett's test of sphericity (IBM, 2021) was also conducted; the value obtained was < .001. A significant chi-square value (χ^2 = 339,012.67) with 36 *df* was obtained, indicating the presence of covariance and a linear relationship between the variables. A scree plot was used to determine the appropriate number of factors visually. Eigenvalues below 0.32 were suppressed (Costello & Osborne, 2005). The eigenvalues for Dimensions 1 and 2 explain 70.48% of the cumulative variance. The remaining dimensions did not reach an eigenvalue of 1 and will not be discussed. PCA was performed using oblique rotation. Significant contributions to factor loadings emerged for six components in Dimension 1 and for three components in Dimension 2 (see Table 4). The first dimension shows high positive factor loadings for HNR 05, HNR 35, HNR 25, *F*0, and SoE and a high negative factor loading for energy. High positive factor loadings were observed in Dimension 2 for H4*–H2* and F2. A high negative factor loading was observed for A3* in Dimension 2.

As race identification based on the speaker's voice was the value of interest, a second PCA was completed using only cases in which the race was incorrectly identified. The KMO value was 0.57, again indicating a marginal suitability for the analysis. Bartlett's test of sphericity was conducted and a significant chi-square value (χ^2 = 74,004.02) with 36 *df* was observed. A scree plot was used to determine the appropriate number of factors visually. Values below 0.32 were suppressed (Costello & Osborne, 2005).

Eigenvalues for Dimensions 1 and 2 explained 69.57% of the cumulative variance, and Dimension 3 accounted for an additional 11.66%, resulting in 81.23% of the variance accounted for in three dimensions. The remaining dimensions did not reach an eigenvalue of 1 and were not included in the discussion. PCA was performed using oblique rotation. Significant contributions to

Table 3. Acoustic variables.

Variable categories	Acoustic variables
Pitch	*F*0
Formant frequencies	F1, F2, F3
Harmonics	H1*–H2*, H2*–H4*, H3*–2k*, H4*–2k*, H2kHz*–H5kHz, H1*
Spectrum slope	H2*, H3*
Inharmonic source/spectral noise	HNR 05, HNR 15, HNR 25, HNR 35, CPP, Energy, SHR,
Other	Dur [vowel duration], A1*, A2*, A3* [first three formants amplitudes], SoE [voicing amplitude]

Note. *F*0 = fundamental frequency; F1 = first formant; F2 = second formant; F3 = third formant; HNR = harmonic-to-noise ratio; CPP = cepstral peak prominence; Energy = the root-mean-square of energy; SHR = subharmonic ratio; Dur = vowel duration; SoE = strength of excitation.
*Denotes corrected value.

Table 4. PCA factor loadings.

Variable	Dimension 1	Dimension 2
HNR 05	0.89	
HNR 25	0.89	
HNR 35	0.89	
*F*0	0.87	
SoE	0.79	
Energy	–0.69	
H4*–2k*		0.91
A3*		–0.76
F2		0.63

Note. PCA = principal component analysis; HNR = harmonic-to-noise ratio; *F*0 = fundamental frequency; SoE = strength of excitation; Energy = the root-mean-square of energy; F2 = second formant.

*Denotes corrected value.

factor loadings emerged for six components in Dimension 1, one component in Dimension 2, and one component in Dimension 3 (see Table 5). The first dimension showed high positive factor loadings for HNR 25, HNR 15, HNR 35, HNR 05, and H1*–H2* with a relatively high negative factor loading for energy. High factor loadings were observed for SHR in Dimension 2 and F2 in Dimension 3.

Discussion

This study explored the interaction of the talker race and the talker regional dialect in a binary (Black or White) race identification task for distinct regional dialects of Southern American English in East and West North Carolina. Two listener groups—one from the Midland dialect region in Indiana and the other from the Southern dialect region in East North Carolina—heard *hVd* words produced by 24 male and female speakers, evenly divided by race and the North Carolina dialect region. The following discussion analyzes our findings and addresses the research questions.

Table 5. PCA factor loadings error responses.

Variable	Dimension 1	Dimension 2	Dimension 3
HNR 25	0.97		
HNR 15	0.96		
HNR 35	0.96		
HNR 05	0.89		
H1*–H2*	0.40		
Energy	–0.39		
SHR		0.98	
F2			0.92

Note. PCA = principal component analysis; HNR = harmonic-to-noise ratio; Energy = the root-mean-square of energy; SHR = subharmonic ratio; F2 = second formant.

*Denotes corrected value.

Q1. Does talker adherence to regional vowel productions that vary from socioethnic/racial expectations affect listener race identification accuracy?

The listeners identified the race of White talkers with greater accuracy than Black talkers and identified East North Carolina talkers more accurately than the West. This result appears to be driven by the listeners' low race identification accuracy of West North Carolina Black talkers. North Carolina West Black talkers were identified with just-above-the-chance accuracy.

The East Black speakers' adherence to the supraregional socioethnic AAS has a positive effect on their race identification compared to the lower race identification for Black talkers from the West. For example, participating in the SVS may be associated with being White in the minds of the listeners in this experiment. Conversely, talker adherence to regional (or supraregional) vowel productions that meet racialized expectations appears to be positively associated with race identification. For example, participating in the AAS and the absence of back vowel fronting may be associated with being Black in the minds of the listeners in this experiment. The distribution of accuracy by race and region supports our hypothesis that listener race identification is affected by Black talker adherence to regional vowel productions that defy listener expectations.

Q2. Does the listener dialect region—a proxy for dialect familiarity—influence race identification accuracy?

Southern listeners were more accurate in race identification compared to (Indiana) Midland listeners. These results suggest that daily exposure to Southern American English is positively associated with race identification accuracy for Southern American speakers in this experiment.

Q3. Does talker gender influence race identification accuracy?

The obtained results revealed no significant differences in listener responses by talker gender. In addition to our analysis of listener race identification of speech stimuli, we also evaluated how the source and filter components of aperiodicity in the speech signal may have influenced listener race categorization. Following the regression analysis, a PCA was completed on the vowel portions of each word. PCA was resolved into two components. Component 1 explained 48.1% of the variance with high positive factor loadings for the normalized harmonics HNR 05, HNR 25, HNR 35, *F*0, and the amplitude of voicing (SoE). A high negative factor loading was observed for the inharmonic (spectral noise) source energy. Previous researchers have observed that White listeners associated lower HNR (more periodic) signals with Black (African American) speakers. The same bias was not found in Black (African American) listeners. In the current study, we found no significant differences in listener performance according to race.

Both the SoE and the Energy are closely related to the intensity of energy in the speech signal. Energy is a quantitative measure of signal amplitude, whereas SoE is a broader construct defined as the slope of the zero-frequency filtered signal around the glottal closure instants. Put plainly, SoE is a measure of the strength of vocal fold closure during the production of voiced speech (Pravena & Govind, 2017). The SoE, derived from the vowel portion of each word in the current data set, is strongly associated with race identification in Component 1 of the PCA. Energy, a measure of perceived average loudness in the speech signal, showed a strong negative association with race identification in Component 1 of the PCA. The SoE, as all the values presented in this text, is simply an extracted numeric measure. Within a general set of physical parameters, the talker can manipulate this measure. The meaningfulness of this manipulation and the derived numeric value may differ among different talker communities. Our finding that SoE is useful in listener evaluations of race points to the need for additional research in this area with Black talkers, including those who use AAE.

Component 2 accounted for an additional 22% of the variance for a total of 70%. The degree of spectral energy attenuation, and tilt at low frequency, below 2 kHz, and the normalized values of F2 were positively correlated with race identification, while A3 showed a strong negative correlation with accurate race identification.

A second PCA was completed using only the error responses to determine whether listeners used the same components for correct and incorrect evaluations of race. Threwee components were extracted from the error response PCA. Component 1 revealed a strong positive association for the same HNR components as the full PCA plus HNR 15 and a smaller positive association for H1*–H2. Smaller HNRs were associated with Black (African American) speakers, as previously described. H1*–H2 is associated with listener perceptions of breathy to modal or creaky to modal voices. Energy was again negatively associated with race identification, although the association was smaller for error-race identification. SHR had a strong and sole association in Component 2, while F2 was the sole factor in Component 3. F2 was identified by Walton and Orlikoff (1994) as a reliable measure of race identification.

The PCA analysis provides information on the voice quality characteristics of listeners used in their perceptual evaluation of the talker race. The factors *F*0, F2, and HNR were previously identified as relevant measures for identifying AAE talkers. In the current experiment, listeners were presented with AAE speakers from two distant dialects. Our hypothesis that listeners would have difficulty accurately categorizing the race of West Black talkers was upheld, and our analysis provided insight into the spectral acoustic factors of talker speech associated with accurate and error-race identification in the presence of unfamiliar regional dialect use by Western AAE speakers. The talker projection of racial and regional identities is signaled by the intersection of multiple acoustic variables. The current results provide insights into some of these aspects.

The results presented here add to the body of evidence on internal variation in AAE. It must be noted the references cited as the foundation of this work illustrate the need to continuously examine the acoustic spectral features of all varieties of U.S. English as some of the findings may be the result of change over time in either or both listener or talker speech behavior and experience. Future researchers should include representative samples of all talkers within a community including Black talkers that use AAE. Listeners from within and from outside the community should be included in perception experiments to ensure universal features are in fact universal across talker and listener groups. This inclusive research method provides for a robust collection and analysis of the speech in each community.

Conclusions

The purpose of this study was to evaluate how talker adherence to regional dialect variation and listener familiarity with regional dialects affect race identification for Black and White talkers. The results showed that when speech tokens of two Black talker groups—one using supraregional productions associated with AAE and the other using regional productions typically associated with regional White U.S. English—are presented to listeners, the Black talkers who use the regional, not the supraregional, productions are frequently miscategorized as White. When talkers, both Black and White, use productions typically associated with their self-identified racial group, they are accurately categorized as self-identified with well-above-chance accuracy.

The results shared in this work contribute to conversation on race identification within and between racial, socioethnic, and socioregional dialect groups. This work shows that, indeed, there is much more to learn about AAE through the modeling of spectral acoustic data. The results suggest that Black talkers who use AAE employ multiple spectral and acoustic factors of voice to signal their regional identity in concert with their race/socioethnic identity and listeners may associate subsegmental aspects of voice with supralinguistic features, including trust, distress, anger, and boredom. Racial identification and these supralinguistic features are both identifiable in the same acoustic spectral features of voice. To our knowledge, this is the first time that these data have been explored. Our results further suggest that race and regional identity are indebted to both the local and supraregional linguistic constructs. Future research should continue investigations of both community internal and external perceptions and the production of talker identity in the communities of practice and listener perceptions of race in concert with other supralinguistic constructs.

Author Contributions

Yolanda Feimster Holt: Conceptualization, Data curation, Formal analysis, Writing – original draft. **Tessa Bent:** Methodology, Writing – review & editing. **Melissa Baese-Berk:** Formal analysis, Validation, Visualization. **Kathrin Rothermich:** Writing – review & editing.

Data Availability Statement

Recordings are not available for sharing. Acoustic spectral data are available upon request to the corresponding author.

Acknowledgments

We would like to thank our research assistants, Jesse Sherman, Danielle Pacholski, and Jodene Jackson, for their assistance with data collection.

References

Anderson, B. (2008). *Migration, accommodation, and language change: Language at the intersection of regional and ethnic identity*. Springer. https://doi.org/10.1057/9780230582422

Andrianopoulos, M. V., Darrow, K. N., & Chen, J. (2001). Multimodal standardization of voice among four multicultural populations: Fundamental frequency and spectral characteristics. *Journal of Voice, 15*(2), 194–219.

Arnson, S., & Farrington, C. (2017). Twentieth Century sound change in Washington DC African American English. *University of Pennsylvania Working Papers in Linguistics*, 23.

Bailey, G., & Thomas, E. (2021). Some aspects of African-American vernacular English phonology. In S. S. Mufwene, J. R. Rickford, G. Bailey, & J. Baugh (Eds.), *African-American English: Structure, history, and use*. Routledge. https://doi.org/10.4324/9781003165330-5

Bernstein, C. (1993). Measuring social causes of phonological variation in Texas. *American Speech, 68*(3), 227–240. https://doi.org/10.2307/455631

Bryden, J. D. (1968). *An acoustic and social dialect analysis of perceptual variables in listener identification and rating of Negro speakers. Final Report*. Office of Education (DHEW), Washington, DC. Bureau of Research. https://eric.ed.gov/?id=ED022186

Calder, J., & King, S. (2022). Whose gendered voices matter?: Race and gender in the articulation of /s/ in Bakersfield, California. *Journal of Sociolinguistics, 26*(5), 604–623. https://doi.org/10.1111/josl.12584

Childs, B., & Mallinson, C. (2004). African American English in Appalachia: Dialect accommodation and substrate influence. *English World-Wide, 25,* 27–50. https://doi.org/10.1075/eww.25.1.03chi

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America, 118*(3), 1661–1676. https://doi.org/10.1121/1.2000774

Coggshall, E., & Becker, K. (2009). The vowel phonologies of African American and White New York City residents. *Publication of the American Dialect Society, 94,* 101–128. https://doi.org/10.1215/-94-1-101

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(7), 1–9. https://doi.org/10.7275/jyj1-4868

Cukor-Avila, P. (2001). The relationship between the evolution of African American and Southern White Vernacular English in the South: 4. Co-existing grammars. In S. L. Lanehart (Ed.), *Sociocultural and historical contexts of African American English* (p. 93–128). John Benjamins Publishing Company. https://doi.org/10.1075/veaw.g27.09cuk

Eberhardt, M. (2009). African American and White vowel systems in Pittsburgh. *Publication of the American Dialect Society, 94*(1), 129–157. https://doi.org/10.1215/-94-1-129

Farrington, C. (2018). Incomplete neutralization in African American English: The case of final consonant voicing. *Language Variation and Change, 30*(3), 361–383. https://doi.org/10.1017/S0954394518000145

Foreman, C. G. (1999). Dialect identification from prosodic cues. *Proceedings of ICPhS99*. https://www.academia.edu/download/80203646/p14_1237.pdf [PDF]

Fridland, V. (2003). 'Tie, tied and tight': The expansion of /ai/ monophthongization in African-American and European-American speech in Memphis, Tennessee. *Journal of Sociolinguistics, 7*(3), 279–298. https://doi.org/10.1111/1467-9481.00225

Fridland, V., & Bartlett, K. (2006). The social and linguistic conditioning of back vowel fronting across ethnic groups in Memphis, Tennessee. *English Language and Linguistics, 10,* 1–22. https://doi.org/10.1017/S1360674305001681

Graff, D., Labov, W., & Harris, W. A. (1986). Testing listeners' reactions to phonological markers of ethnic identity: A new method for sociolinguistic research. *Diversity and Diachrony, 53,* 45–58.

Green, L. J. (2002). *African American English: A linguistic introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511800306

Hawkins, F. D. (1992). *Speaker ethnic identification: The roles of speech sample, fundamental frequency, speaker, and listener variations* [Doctoral dissertation, University of Maryland]. https://www.proquest.com/openview/935a3747df994f9f11bd0acf0f17ab30/1?pq-origsite=gscholar&cbl=18750&diss=y

Holliday, N. R., & Jaggers, Z. (2015). Influence of suprasegmental features on perceived ethnicity of American politicians. *ICPhS*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0354.pdf [PDF]

Holt, Y. F.(2011). *A cross generational dialect study in Western North Carolina* [Doctoral dissertation, The Ohio State University]. https://rave.ohiolink.edu/etdc/view?acc_num=osu1299037925

Holt, Y. F. (2018). Mechanisms of vowel variation in African American English. *Journal of Speech, Language, and Hearing Research, 61*(2), 197–209. https://doi.org/10.1044/2017_JSLHR-S-16-0375

Holt, Y. F., & Ellis, C. (2018). African American women's speech: Vowel inherent spectral change. *Acoustical Science and Technology, 39,* 160–162. https://doi.org/10.1250/ast.39.160

Holt, Y. F., Jacewicz, E., & Fox, R. A. (2015). Variation in vowel duration among Southern African American English speakers. *American Journal of Speech-Language Pathology, 24*(3), 460–469. https://doi.org/10.1044/2015_AJSLP-14-0186

Holt, Y. F., & Rangarathnam, B. (2018). *F*0 declination and reset in read speech of African American and White American women. *Speech Communication, 97,* 43–50. https://doi.org/10.1016/j.specom.2018.01.001

Hudson, A. I., & Holbrook, A. (1981). A study of the reading fundamental vocal frequency of young Black adults. *Journal of Speech and Hearing Research, 24*(2), 197–200. https://doi.org/10.1044/jshr.2402.197

IBM. (2021, November 9). *KMO and Bartlett's Test*. Retrieved June 24, 2024, from https://www.ibm.com/docs/en/spss-statistics/28.0.0?topic=detection-kmo-bartletts-test

Irons, T. (2007). On thestatusoflow back vowels in Kentucky English: More evidence of merger. *Language Variation and Change, 19*(2), 137–180. https://www.cambridge.org/core/journals/language-variation-and-change/article/abs/on-the-status-of-low-back-vowels-in-kentucky-english-more-evidence-of-merger/068EA2C863270C13F5C6E01CB9DB8701

King, S. (2016). On negotiating racial and regional identities: Vocalic variation among African Americans in Bakersfield, California. *University of Pennsylvania Working Papers in Linguistics, 22*(2), 12.

King, S., & Calder, J. (2020). 4. The nature of BOOT fronting among African Americans in Bakersfield, California. *Publication of the American Dialect Society, 105*(1), 64–78. https://doi.org/10.1215/00031283-8820620

Labov, W. (2006). *The dialects of North American English*. Mouton de Gruyter. https://doi.org/10.1515/9783110167467.d.119

Labov, W., Ash, S., & Boberg, C. (2008). *The Atlas of North American English: Phonetics, phonology and sound change*. De Gruyter Mouton. https://doi.org/10.1515/9783110167467

Labov, W., Fisher, S., Gylfadottír, D., Henderson, A., & Sneller, B. (2016). Competing systems in Philadelphia phonology. *Language Variation and Change, 28*(3), 273–305. https://doi.org/10.1017/S0954394516000132

Lanehart, S. L. (2015). *The Oxford handbook of African American language*. Oxford University Press.

Lass, N. J., Almerino, C. A., Jordan, L. F., & Walsh, J. M. (1980). The effect of filtered speech on speaker race and sex identifications. *Journal of Phonetics, 8*(1), 101–112. https://doi.org/10.1016/S0095-4470(19)31445-7

Lass, N. J., Mertz, P. J., & Kimmel, K. L. (1978). The effect of temporal speech alterations on speaker race and sex identifications. *Language and Speech, 21*(3), 279–290. https://doi.org/10.1177/002383097802100306

Lass, N. J., Tecca, J. E., Mancuso, R. A., & Black, W. I. (1979). The effect of phonetic complexity on speaker race and sex identifications. *Journal of Phonetics, 7*(2), 105–118. https://doi.org/10.1016/S0095-4470(19)31044-7

Lee, S. (2011, October). *High and mid back vowel fronting in Washington, DC, Metropolitan Area* [Paper presentation]. 40th Annual Conference on New Ways of Analyzing Variation (NWAV 40), Washington, DC.

Lee, S. (2018). Patterns of the mainstream sound change in a liminal region: Low back merger in Washington DC. *Journal of English Linguistics, 46*(4), 267–292. https://doi.org/10.1177/0075424218788923

Majors, T. (2005). Low back vowel merger in Missouri speech: Acoustic description and explanation. *American Speech, 80*(2), 165–179. https://doi.org/10.1215/00031283-80-2-165

Mallinson, C., & Childs, B. (2004). The intersection of regional and ethnic identity: African American English in Appalachia. *Journal of Appalachian Studies, 10*(1/2), 129–142.

Mallinson, C., & Childs, B. (2007). Communities of practice in sociolinguistic description: Analyzing language and identity practices among Black women in Appalachia. *Gender and Language, 1*(2), 173–206. https://doi.org/10.1558/genl.v1i2.173

McLarty, J. (2018). African American language and European American English intonation variation over time in the American South. *American Speech, 93*(1), 32–78. https://doi.org/10.1215/00031283-6904032

Mengesha, Z. (2022). *A social meaning perspective on vowel trajectories: The FEEL-FILL merger among African Americans.*

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

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition, 114*(1), 42–55. https://doi.org/10.1016/j.cognition.2009.08.012

Pravena, D., & Govind, D. (2017). Development of simulated emotion speech database for excitation source analysis. *International Journal of Speech Technology, 20,* 327–338. https://doi.org/10.1007/s10772-017-9407-3

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

Risdal, M. L., & Kohn, M. E. (2014). Ethnolectal and generational differences in vowel trajectories: Evidence from African American English and the Southern Vowel System. *University of Pennsylvania Working Papers in Linguistics, 20,* 139–148.

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

Spears, A. K. (1988). Black American English. In J. B. Cole (Ed.), *Anthropology for the nineties: Introductory readings* (pp. 96–113). Free Press.

Thomas, E. R. (1989). The implications of /o/ fronting in Wilmington, North Carolina. *American Speech, 64*(4), 327–333. https://doi.org/10.2307/455724

Thomas, E. R. (2002). Sociophonetic applications of speech perception experiments. *American Speech, 77*(2), 115–147. https://doi.org/10.1215/00031283-77-2-115

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., & Bailey, G. (2015). Segmental phonology of African American English. In S. Lanehart (Ed.), *The Oxford handbook of African American language* (pp. 403–419). Oxford University Press.

Thomas, E. R., & Coggshall, E. L. C. (2007). Comparing phonetic characteristics of African American and European American Speech. *Linguistica Atlantica, 27,* 112–116. https://journals.lib.unb.ca/index.php/la/article/view/32492

Thomas, E. R., & Lanehart, S. L. (2015). Prosodic features of African American English. In S. L. Lanehart (Ed.), *The Oxford handbook of African American language* (pp. 420–438). Oxford University Press.

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

Vicenik, C., Lin, S., Keating, P., & Shue, Y.-L. (2024). *VoiceSauce: A program for voice analysis* [Computer software].

Walton, J. H., & Orlikoff, R. F. (1994). Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech, Language, and Hearing Research, 37*(4), 738–745. https://doi.org/10.1044/jshr.3704.738

Wassink, A. B. (2015). Sociolinguistic patterns in Seattle English. *Language Variation and Change, 27*(1), 31–58. https://doi.org/10.1017/S0954394514000234

Wolfram, W. (2007). Sociolinguistic folklore in the study of African American English. *Language and Linguistics Compass, 1*(4), 292–313. https://doi.org/10.1111/j.1749-818X.2007.00016.x

Wolfram, W. (2019). African-American English. In C. L. Nelson, Z. G. Proshina, & D. R. Davis (Eds.), *The handbook of world Englishes* (2nd ed., pp. 314–330). Wiley. https://doi.org/10.1002/9781119147282.ch18

Wolfram, W., & Fasold, R. W. (1974). *The study of social dialects in American English*. Prentice-Hall. https://covers.openlibrary.org/b/id/10511265-L.jpg

Wolfram, W., & Thomas, E. (2002). *The development of African American English*. Wiley. https://doi.org/10.1002/9780470690178.ch7

Wolfram, W., Thomas, E. R., & Green, E. W. (2000). The regional context of earlier African American speech: Evidence for reconstructing the development of AAVE. *Language in Society, 29*(3), 315–355. https://doi.org/10.1017/S0047404500003018

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

Xue, S. A., & Fucci, D. (2000). Effects of race and sex on acoustic features of voice analysis. *Perceptual and Motor Skills, 91*(3, Pt. 1), 951–958. https://doi.org/10.2466/pms.2000.91.3.951