Published in final edited form as:
Cognition. 2010 January ; 114(1): 42–55. doi:10.1016/j.cognition.2009.08.012.

Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices

Tyler K. Perrachione[1,4,*], **Joan Y. Chiao**[2,4,5], and **Patrick C.M. Wong**[3,5,6]

[1]Department of Linguistics, Northwestern University, Evanston, Illinois 60208, USA

[2]Department of Psychology, Northwestern University, Evanston, Illinois 60208, USA

[3]Roxelyn & Richard Pepper Department of Communication Sciences & Disorders, Northwestern University, Evanston, Illinois 60208, USA

[4]Cognitive Science Program, Northwestern University, Evanston, Illinois 60208, USA

[5]Northwestern University Interdepartmental Neuroscience Program, Northwestern University, Evanston, Illinois 60208, USA

[6]Department of Otolaryngology – Head & Neck Surgery, Northwestern University, Evanston, Illinois 60208, USA

Abstract

The own-race bias in memory for faces has been a rich source of empirical work on the mechanisms of person perception. This effect is thought to arise because the face-perception system differentially encodes the relevant structural dimensions of features and their configuration based on experiences with different groups of faces. However, the effects of sociocultural experiences on person perception abilities in other identity-conveying modalities like audition have not been explored. Investigating an own-race bias in the auditory domain provides a unique opportunity for studying whether person identification is a modality-independent construct and how it is sensitive to asymmetric cultural experiences. Here we show that an own-race bias in talker identification arises from asymmetric experience with different spoken dialects. When listeners categorized voices by race (White or Black), a subset of the Black voices were categorized as sounding White, while the opposite case was unattested. Acoustic analyses indicated listeners' perceptions about race were consistent with differences in specific phonetic and phonological features. In a subsequent person-identification experiment, the Black voices initially categorized as sounding White elicited an own-race bias from White listeners, but not from Black listeners. These effects are inconsistent with person-perception models that strictly analogize faces and voices based on recognition from only structural features. Our results demonstrate that asymmetric exposure to spoken dialect, independent from talkers' physical characteristics, affects auditory perceptual expertise for talker identification. Person perception thus additionally relies on socioculturally-acquired dynamic information, which may be represented by different mechanisms in different sensory modalities.

Corresponding Author: Patrick Wong, Department of Communication Sciences & Disorders, 2240 Campus Drive, Evanston, IL 60208, Ph: 847.491.2416, Fx: 847.491.2429, pwong@northwestern.edu.
*Current Address: Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

1. Introduction

The human experience is substantially a social one, a fact reflected in the functional configuration of our nervous system. From cortex dedicated to the perception of faces (Kanwisher, McDermott & Chun, 1997), voices (Belin et al., 2000) and bodies (Downing, 2001), we are uniquely adapted to think about other people. Our social and cultural environment contributes to tuning the cognitive and perceptual functions of our nervous system (Chiao & Ambady, 2007; Chiao et al., 2008; Wong et al., 2007; Wong et al., 2004). We begin to organize the world and individuals in it along socially-relevant dimensions in the first few months of infancy (Pascalis et al., 2005), during which exposure to different types of faces gives rise to an own-race bias in face perception by adulthood (Hayward, Rhodes, & Schwaninger, 2008) – an effect reflected in neural responses to own- and other-race faces (Golby et al., 2001). In the auditory modality we become sensitive to the specific sounds of our own language (Kuhl et al., 1992), as well as the particular manners of speech of those closest to us (Kinzler, Dupoux, & Spelke, 2007). Because the human experience is not singular, individuals' abilities and expertise differ considerably and are influenced by *asymmetric cultural experiences*, meaning the quantity or quality of some experiences exceeds that of others. Currently, the extent to which culture influences person perception in the auditory domain remains unknown. Understanding the role of cultural experience in talker identification abilities will not only serve to more fully describe the mechanisms of auditory person perception, it will also help reveal the overarching roles of experience in person-perception abilities, including the ways in which its role in shaping visual and auditory expertise might differ. In two experiments, we investigate how asymmetric cultural experiences affect perceptual expertise for voices of different backgrounds. In Experiment 1 (Voice-Race Categorization), adult African-American ("Black") and Caucasian-American ("White") listeners attended a number of voices and indicated for each token whether they believed a White or Black individual was speaking. Acoustic analyses of salient features of African American English were conducted to determine whether the presence of specific phonetic and phonological (dialectal) features were predictive of listeners' perceptions of race from voice. In Experiment 2, another group of participants learned to identify by name the individual voices of a number of Black and White talkers.

The own-race bias in memory for faces has been a rich source of empirical work on the mechanisms of person perception (Meissner & Brigham, 2001). This effect is thought to arise because the face-perception system differentially encodes the relevant dimensions of structural features and their configuration based on asymmetric exposure to groups of faces (Hayward, Rhodes, & Schwaninger, 2008). Current paradigms of person perception strictly analogize voice perception (Belin, Fecteau, & Bédard, 2004; Campanella & Belin, 2007) to face perception (Bruce & Young, 1986), describing voice perception as a process that exclusively computes differences in vocal structure (e.g. vocal tract length, oral cavity volume, fundamental frequency dynamic range). Such a structure-only model predicts the perceptual categorization and identification of voices will be based exclusively on structural features of oropharyngeal anatomy. Putative covariance with differences in the other physical features that are canonically indicative of an individual's race would give rise to the perception of race from voice. Thus, structure-only models predict that like faces, listeners will exhibit an own-race bias only for voices of the same race as themselves. We call the predictions of this model the *Anatomical-Race Hypothesis* because it is based on the presupposition that, directly analogous to face perception, asymmetric experience with physical (structural) differences between racial groups will be both necessary and sufficient to explain an own-race bias in voice perception.

Despite the current focus on a structure-only model of person perception from voice in the contemporary literature, there are many compelling reasons to doubt its explanatory adequacy with regards to the full range of auditory perceptual abilities humans demonstrate for voices.

First, in contrast to physical feature-configuration properties of Black and White faces, the physical features of their vocal tracts (e.g., volume and length of the oral and pharyngeal cavities, frequencies of the first three formants of the steady-state vowel /a/) do not appear to differ significantly between these two racial groups (Xue, Hao, & Mayo, 2006). That is, a substantial amount of variability exists along any of the dimensions of these features, and the range of this variability is shared among members of both racial groups. Second, there is much evidence that variable information in a talker's utterances, such as the phonetics and other nonlinguistic idiosyncratic manners of speech, is not only sufficient for talker identification (Remez, Fellowes, & Rubin, 1997), but that this information is in fact an important component of natural talker identification (Perrachione & Wong, 2007). Third, individuals of the same race develop different manners of speech (Evans & Iverson, 2004), whereas individuals of different races may be indistinguishable in dialect (Thomas & Reaser, 2004). Fourth, the voice of a single individual talker may be differentially categorized by naïve listeners as being White, Black, or Hispanic, depending on the dialect adopted by the speaker (Purnell, Idsardi, & Baugh, 1999). Taken together, such results suggest that unlike faces, an own-race bias for voices is unlikely to result from asymmetric exposure to the structural features of vocal anatomy, given that such features do not exhibit significant differences across racial groups. Instead, an own-race bias for voices is more likely to arise from asymmetric exposure to the dynamic, culturally-acquired features of spoken language, which in many cases do covary with racial group.

The alternative hypothesis, which we call the *Dialectal-Race Hypothesis*, is based on a recently proposed model (Perrachione & Wong, 2007) that allows for both physical and socially-acquired features to contribute directly to person perception. This model is schematized in Fig. 1. Such a model predicts that the categorization of voices by race largely relies on knowledge of socially-acquired dialectal idiosyncrasies stereotypically associated with members of that race (Thomas & Reaser, 2004; Purnell, Idsardi, & Baugh, 1999). An own-race bias in voice perception is thus likely to occur because listeners have asymmetric exposure to different spoken dialects, and listeners will therefore show an advantage for identifying voices not only of their own race, but also voices of another race that share dialectal features of the listener. Moreover, listeners should not exhibit an own-race bias for voices of their own race with whom they do not share the same socially-acquired dialectal features. Based on the results of two experiments on voice-race categorization and individual talker identification, we demonstrate that structure-only models are incommensurate with the full range of auditory features used by humans in person identification, and that a model that integrates socially-acquired features provides a more complete concept of what mechanisms may underlie our person-identification abilities.

2. Experiment 1: Voice-Race Categorization

Self-identified Black and White participants listened to recordings of voices reading sentences and indicated on each trial whether they thought a Black or White individual was speaking. This experiment was designed to assess whether listeners' perception of race is based on the structural features of voices (per the Anatomical-Race Hypothesis) or on race-independent features of spoken dialectal (per the Dialectal-Race Hypothesis). Subsequent acoustic and phonological analyses provided further tests of these hypotheses by examining whether listeners' perceptions about race were consistent with differences in specific phonetic features.

2.1 Stimuli

Twelve Black males and twelve White males were digitally recorded reading 10 sentences taken from the Harvard Sentences (Institute of Electrical and Electronics Engineers [IEEE], 1969), a collection of phonetically-balanced sentences. These sentences were used in our previous talker-identification experiment (Perrachione & Wong, 2007) and are reproduced here in Appendix A. No speaker who produced the stimuli took part in the subsequent Voice-Race

Categorization experiment. To control for non-race-specific factors affecting acoustic or phonetic differences in voice, subjects were matched between the two racial groups for physical factors thought to be conveyed by vocal information (Evans, Neave, & Wakelin, 2006; Bruckert et al., 2006; cf. Lass & Brown, 1978). Information about the talkers' characteristics is provided in Appendix B. Talkers between the two groups did not differ for height [$t(22) = -0.331$, $p = 0.743$], weight [$t(22) = 1.668$, $p = 0.110$], or age [$t(22) = 0.121$, $p = 0.905$; all independent sample, two-tailed]. All talkers were native speakers of American English, and grew up in the United States of America.

In a sound-attenuated chamber, talkers were instructed to read the sentences as though having a conversation with a friend and were digitally recorded at 44.1 kHz. Recordings were transferred to a PC where they were cut into individual stimuli, resampled to 22.05 kHz using Praat (Boersma & Weenink, 2008) for compatibility with the stimulus-delivery software, and normalized for RMS amplitude to 70 dB SPL using Level16 to insure homogenous recording quality across speakers and sentences. Note that by normalizing stimuli for RMS amplitude, absolute loudness was largely removed as a cue to talker identity.

2.2 Participants

Additional listeners of both races participated in categorizing our stimuli by race. The Black listener group consisted of 10 individuals (all female) ages 18-22 years (M = 19.8), all of whom self-identified as Black and came from entirely African-American families. The White listener group consisted of 10 individuals (7 females) ages 18-24 years (M = 20.2), all of whom self-identified as White and came from entirely Caucasian-American families. All participants in this study (Experiments 1 & 2) were students or staff at Northwestern University, members of the local community, or their friends and family. All participants reported having normal speech and hearing and being free from psychological or neurological deficits. All participants (talkers and listeners) gave informed written consent overseen by the Northwestern University Institutional Review Board, and received either a cash payment or class credit as compensation.

2.3 Procedure

Participants in the categorization experiment heard all 24 voices reading each of the 10 sentences twice, with stimuli randomized by speaker and sentence (24 voices × 10 sentences × 2 repetitions = 480 trials). Stimuli were presented binaurally over headphones in a sound-attenuated chamber; written instructions accompanying the experiment were presented to participants on a computer monitor. After listening to a speaker read a sentence, participants indicated whether they thought the speaker was Black or White by pressing either of two buttons on a response box. Listeners were told their responses would help us select stimuli for a future experiment, and they should make their determinations about racial identity based on their personal experiences. The task was self-paced and generally took participants approximately 35-40 minutes to complete. Participant accuracy was averaged by talker (Fig. 2) and analyzed with a 2×2 repeated measures ANOVA for main effects and interactions of talker race or listener race. Data were also analyzed using signal detection theory for sensitivity and response bias.

2.4 Results and Discussion

Although listeners were overall fairly accurate at this task, listeners' categorization of some of the voices deviated substantially from the talkers' self-reported race. A 2×2 repeated measures ANOVA indicated White voices were overall more accurately categorized than Black voices [$F(1,18) = 58.147$, $p < 0.001$]. By examining listeners' categorization accuracy on each voice individually (Fig 2a), we found that certain Black voices were nearly always categorized as Black, whereas other Black voices were frequently miscategorized as sounding White. Meanwhile, the White voices were only rarely miscategorized as sounding Black. That some

voices were most frequently categorized as members of the other racial group indicates that auditorily-perceived race is not always consistent with visually-perceived or self-reported race. Overall accuracy was not affected by listener race [$F(1,18) = 0.000$, $p = 0.991$], nor was there an interaction between listener and talker race [$F(1,18) = 0.125$, $p = 0.728$].

To further investigate any distinction between the two groups' performance on the categorization task, participants' responses were also analyzed using signal detection theory. We arbitrarily chose Black voices to be the signal, such that the correct categorization of a Black voice is a “hit” and the categorization of a White voice as Black was a “false alarm” and so on. (Choosing White voices as the signal produces equivalent results.) As shown in Fig. 2b, Black listeners were on average more sensitive to racial voice information than White listeners (mean $d'_{Black} = 2.339$, mean $d'_{White} = 1.931$, $t(18) = 2.105$, $p < 0.05$). This difference in sensitivity (in light of the lack of a main effect of listener race on overall accuracy) arose due to a difference in response bias, as shown in Fig. 2c. Although both listener groups were conservative in their responses (that is, both had a tendency to say the “Black voice” signal was absent), the Black listeners as a group were significantly more conservative (less likely to respond “Black”) than the White listener group (mean $\ln \beta_{Black} = 1.584$, mean $\ln \beta_{White} = 0.725$, $t(18) = 2.284$, $p < 0.035$). The bias towards categorizing a voice as “White” may be a statistical reflection of the overall proportion of Black or White individuals that our participants encountered in their daily lives. That is, given a racially ambiguous voice, listeners might be biased towards responding “White” given the prevalence of white individuals in some relevant, experience-related basis set. Alternately, this bias suggests there are likely to be specific cues necessary to evoke a percept of that race. The selection of “White” as the default response can thus help inform an understanding of the acoustic-phonetic features that are necessary or sufficient for a voice to be perceived as a member of a minority group. This account is supported by the greater bias demonstrated by Black listeners against categorizing voices as Black – a result not predicted by an environment-driven bias. Future hypothesis-driven research is needed to clarify the social-psychological issues surrounding auditory features and group membership.

In Experiment 1, participants were able to infer racial identity from an individual's voice, which is consistent with both our everyday experiences and previous empirical studies (Thomas & Reaser, 2004; Purnell, Idsardi, & Baugh, 1999). The categorization responses from listeners of both races distinguished three groups of voices: Black voices that “sound Black” (Group 1), White voices that “sound White” (Group 2), and another group of Black voices that were most frequently perceived as “sounding White” (Group 3). We confirmed that listeners did in fact categorize these latter voices as White more often than chance [$t(19) = -2.088$, $p = 0.025$]. White voices were only rarely miscategorized as sounding Black, leaving us without a fourth group exemplifying the reverse situation of Group 3. These results support the underlying premise of the Dialectal-Race Hypothesis that talkers are distinguished based on features of their speech independent from the physical (visual) or structural (vocal) features on which their self-described racial identity may be based. That is, if racial identity were conveyed by invariant structural information in the voice, we should have seen completely binary classification of the voices. That a voice can sometimes sound like a member of one racial group and other times like a member of a different racial group suggests racial identity is largely computed over variable cues, such as features of the spoken dialect.

In our analysis, all talkers read the same sentences, which removed any morphosyntactic cues to racial identity, leaving only phonetics and other markers specific to speech. Also of note is the fact that, despite matching the two racial groups for physical features commonly shown to affect vocal tract dimensions such as age, height, and weight (Evans, Neave, & Wakelin, 2006; Bruckert et al., 2006; see Appendix B), none of the White talkers was frequently categorized as sounding Black. If the perception of racial group membership had arisen based

on particular structural dimensions, the physical variability built into our sample should have resulted in some of the White voices frequently categorized as sounding Black, just as some Black voices were frequently categorized as sounding White. The lack of such a group strongly suggests there is no “White-sounding” or “Black-sounding” vocal prototype based on physical dimensions alone. Having ruled out the ability of anatomical differences to explain listeners' perceptions about race, we next investigated whether listeners' perceptions were consistent with socially-acquired phonetic features of talkers' dialects.

3. Acoustic-Phonetic and Phonological Features Associated with Listener-Categorized Race

Listeners' categorization results from Experiment 1 strongly suggested the perception of race from auditory information was based on differences in spoken dialect rather than differences in vocal anatomy. To verify this conclusion, we examined whether specific dialectal features were sufficient to distinguish Group 1 from Group 2. Moreover, if these differences in dialectal features are indeed the perceptual basis for categorizing talker race, Group 3 should bear more similarity to Group 2 on these features than Group 1. Numerous sociophonetic studies have examined differences in phonetic and phonological features across dialects of American English (e.g. Clopper & Pisoni, 2004), many of which focus on dialectal features that vary by racial group. Such studies and the features they identify have been comprehensively reviewed by Thomas and Reaser (2004). To quantify dialectal differences across the three talker groups at the phonological level, we first assessed the groupwise distribution of major phonological features common to African American English dialects. We also investigated two prominent acoustic-phonetic features – cardinal vowel space and consonantal voice-onset time – shown in the sociophonetic literature to be differentially represented in Black and White dialects of American English to determine whether dialectal differences in articulatory features alone were consistent with the perceptual delineation of the three talker groups from Experiment 1.

3.1 Distribution of Common Phonological Features of African-American English Dialects

There are many systematic phonological and morphosyntactic features of African-American English dialects (elsewhere sometimes referred to as African American Vernacular English) which make it perceptually distinct from “Standard American English” to both naïve listeners and linguists (Green, 2002). Talkers in our experiment read predetermined sentences as stimuli, such that morphosyntactic features distinguishing African-American English were not available to our listeners. These stimuli sentences, however, did afford ample opportunity for talkers to express the phonological features of African-American English, even though they had not been designed specifically to elicit those features. Here we investigated the prevalence of these phonological features across the three talker groups, testing the hypothesis that Group 1 would exhibit more frequent use of such features than Groups 2 or 3.

3.1.1 Acoustic Measurements—From accounts of the major phonological features of African-American English (Pollock & Meredith, 2001; Craig et al., 2003), we assembled a list of 12 features that could occur in our stimuli, of which nine were attested at least once: Reduction of voiced final consonant clusters, deletion of /r/ in clusters before rounded vowels, deletion or devoicing of final obstruents, stopping or labialization of interdental fricatives, vocalization of postvocalic and syllabic /l/, derhotacization or deletion of vocalic and postvocalic /r/, and monophthongization or /ai/. Based on the canonical phonological representation of each stimulus sentence, we determined there were 50 points where one of these features could be realized.

For each of the sentences in our stimulus set, we determined whether one of these phonological features was realized at each possible point. Specific acoustic-phonetic criteria were

established for each rule to avoid listener expectation effects. The number of times each feature occurred was summed within each talker group, and the total number of feature occurrences was summed for each individual talker.

3.1.2 Results—The groupwise incidence of these phonological features is illustrated in Fig. 3, and was submitted to a univariate ANOVA where the dependent variable was the number of African-American English phonological features observed and the two independent variables were the talkers' listener-perceived and self-identified racial identity. The incidence of these phonological features was not explained by talkers' self-identified race [$F(1,12) = 0.092$, $p = 0.767$]. However, there was a significant difference in the incidence of these features based on listener-categorized racial identity [$F(1,12) = 24.939$, $p < 0.0004$]. Group 1, which listeners categorized as Black, had a significantly higher incidence of phonological features typical of African-American English than either Group 2 or 3, which listeners categorized as sounding White. The incidence of these features did not differ between Groups 2 and 3.

We also investigated whether the types of these phonological features were differentially attested across the three groups. The frequencies of phonological features were compared pairwise between the three groups; features attested fewer than five times were excluded to meet the assumptions of the statistical test, and corrections were made for multiple comparisons. The types of features exhibited by Group 1 differed significantly from both Group 2 [$\chi^2(4) = 19.38$, $p < 0.003$] and Group 3 [$\chi^2(4) = 18.34$, $p < 0.005$]. Groups 2 and 3, however, again did not differ from one another [$\chi^2(3) = 8.71$, $p = 0.136$]. Thus, the types of African-American English phonological features exhibited by Group 1, which listeners categorized as sound Black, differed from Groups 2 and 3, which were categorized as sounding White. In particular, Group 1 exhibited more frequent deletion of final consonants and stopping of interdental fricatives than Groups 2 or 3, and more frequently vocalized postvocalic /l/ than Group 2.

3.2 Vowel Quality

Vowels are described by the concentration of spectral energy in the first and second formant (F1 and F2). Differences in vowel quality often serve to distinguish various dialects of a language (e.g. Evans & Iverson, 2004), and their differences among racial dialects have been studied extensively (Thomas, 2001; Thomas & Reaser, 2004). Here we investigated the vowel space of the three talker groups as defined by the cardinal vowels /a/, /i/, and /u/ (as in “pot,” “heat,” and “booth.”)

3.2.1 Acoustic Measurements—Tokens from the test sentences in Experiment 1 were used in this analysis. The following tokens from each talker were used: for /a/, “rod,” “pot,” “soft,” and “across”; for /i/, “feet,” “tea,” “evening,” “heat,” “breeze,” “sea,” and “fifty”; and for /u/, “huge,” “through,” and “booth.” Vowels were measured through their longest steady-state portion beginning at least two periods of phonation after the preceding phoneme (to avoid coarticulatory artifacts) and stopping at least two periods before either the coda consonant, the onset of creaky voice, or the end of phonation. The mean values of F1 and F2 in the region described above were determined using the formant tracker implemented in Praat.

3.2.2 Results—The cardinal vowel spaces for the three talker groups are shown in Fig. 4. The vowel measurements were submitted to multivariate ANOVA, with F1 and F2 as the dependent variables, and Vowel, Listener-Categorized Race, and Self-Described Race as independent variables. The results indicated that vowel measurements differed significantly by listener-categorized race for both F1 [$F(1,201) = 26.490$, $p < 6.3 \times 10^{-7}$] and F2 [$F(1,201) = 6.544$, $p < 0.011$]. On the other hand, differences associated with talkers' self-described racial identity were not significant [F1: $F(1,201) = 0.179$, $p = 0.673$; F2: $F(1,201) = 0.062$, $p = 0.803$],

consistent with earlier measurements by Xue, Hao, & Mayo, (2006). There was an interaction between vowel and listener-categorized race for F2 [$F(2,201) = 8.618$, $p < 0.0003$], indicating a substantially larger magnitude in the difference between Group1 and Groups 2 and 3 for /u/ (see Fig. 4). There were no interaction effects with talkers' self-described racial identity.

The phonetic implications of these statistics are clearly evident in Fig. 4. The two talker groups categorized by listeners as sounding White (Groups 2 and 3) have virtually identical means and distributions for F1 and F2 of all cardinal vowels, whereas the group categorized as sounding Black (Group 1) exhibits a consistently lower F1 than the other groups, in addition to differences in F2 that vary by vowel. In particular, Groups 2 and 3 exhibited consistent fronting of the vowel /u/ relative to Group 1 – a feature typical of many White dialects (Thomas, 2001).

3.3 Voice-Onset Time

Sounds like /b/ and /p/ are distinguished by when phonation begins relative to when the consonantal constriction is released. In addition to phonemic distinctions in voice-onset time (VOT) within a language, this feature can also vary between languages and dialects. Ryalls, Zipprer, and Baldauff (1997) found that African-American talkers displayed significantly more prevoicing (a negative VOT) for voiced stop consonants than Caucasian-American talkers. Here we ask whether this difference in VOT also distinguishes our three talker groups.

3.3.1 Acoustic Measurements—For all 5 talkers in each of Groups 1, 2, and 3, VOTs were measured from word-initial voiced stop consonants in the sentences heard by the listeners in Experiment 1 and included the following tokens: “boy,” “ball,” “back,” “broke,” “breeze,” “booth,” and “bonds.” Acoustic measurements were carried out in Praat, following the method described in Ryalls, Zipprer, and Baldauff (1997). VOT was defined as the time between the onset of the burst and the onset of phonation (identified in both the spectrogram and waveform). Negative VOTs were determined as the time between the onset of phonation and the consonantal release, and were only included if voicing continued through the burst.

3.3.2 Results—The values of VOTs were not normally distributed, as determined by Shapiro-Wilk tests for normality, and so comparisons between groups were made using the Mann-Whitney U statistic. Consistent with the report of Ryalls and colleagues (1997), Group 1 exhibited significantly more prevoicing than Group 2 ($W = 418.5$, $p < 0.012$, one-tailed). Group 2 and Group 3, whom listeners most frequently categorized as sounding White, did not differ in the amount of prevoicing ($W = 541$, $p = 0.202$, one-tailed), whereas there was a reliable trend for Group 1 to exhibit more prevoicing than Group 3 ($W = 478$, $p < 0.058$, one-tailed), despite these groups' shared self-described racial identity. In sum, the dialectal features of voicing were shared between Groups 2 and 3, while distinguishing both from Group 1.

3.4 Discussion

We analyzed a set of phonological features and two phonetic features previously shown to vary by race-associated dialect. Here we observed that not only do these features distinguish Group 1 from Group 2, they moreover show that such shared phonetic and phonological features are consistent with listeners' perceptions of Group 3 talkers as sounding White. Crucially, at every opportunity, the associations between the three groups emerged consistent with listener-categorized race; Groups 2 and 3 patterned together distinct from Group 1. Nowhere was there evidence of a distinction by self-described racial identity; Groups 1 and 3 never patterned together distinct from Group 2.

In natural, running speech, such as our sentence stimuli, a variety of dialectal cues are rapidly available to listeners, from which they are able to form a robust impression of a talker's race

based on the presence or absence of stereotypic phonetic and phonological features. Our results are thus directly consistent with earlier work showing that listeners' accuracy at categorizing race from voice increases as a function of the phonetic complexity of the stimulus set (Lass, Tecca, Mancuso, & Black, 1979). Figure 5 provides an example of how these features combine to clearly contrast Black-sounding from White-sounding productions. In a Group 1 production of “booth,” the presence of both prevoicing and a low F2 /u/ are clearly evident, compared to a Group 2 production in which there is a slight positive VOT and a relatively higher (fronted) /u/.

It is especially compelling that none of the features for which we found group differences can be accounted for by specific differences in anatomy. The phonological features investigated are the result of dynamic, experience-dependent processes that apply on-line during speech production, and the incidence of which is associated with differences in dialect, not physiology. Voice-onset time is a canonical feature associated with specific phonemes, which varies across dialects and languages. Although vowel quality can differ based on anatomical differences between talkers (e.g. males vs. females), differences in dialect also produce reliable individual differences in vowel quality (Clopper, Pisoni, and de Jong, 2005). Between the results of Experiment 1 and these acoustic-phonetic analyses, we can be confident that specific features of spoken dialect are sufficient to provide a robust percept of race from auditory information (Purnell, Idsardi, and Baugh, 1999).

It bears noting that our phonetic and phonological analyses of the recordings were decidedly *post hoc* – the stimuli were not specifically designed to elicit the phonological features of African American English, nor did we assess racial categorization under parametric variation of these features. Although we showed that differences in these contrasts were consistent with listeners' perceptions of race from voice – and previous work has indicated listeners are indeed sensitive to individual variability in such contrasts (e.g. Allen & Miller, 2004) – future work is still necessary to show a causal relationship between these features and perception of race from voice.

4. Experiment 2: Individual Talker Identification

The categorization data indicate perceived race is based on features of spoken dialect, not anatomical features, but they do not speak to whether the canonical own-race bias exists in memory for voices, nor whether asymmetric experience with differences in vocal structure or dialectal features across racial groups are the underlying source of this effect. To investigate these questions, we ran an individual talker identification experiment, in which Black and White listeners learned to identify individual talkers in the three different groups by name. If there is an own-race bias in memory for voices, Black listeners should be more accurate at identifying Group 1 voices, and White listeners should be more accurate at identifying Group 2 voices.

However, the two hypotheses discussed earlier make very different predictions about the basis for this memory bias. The Anatomical-Race Hypothesis holds that asymmetric experience with differences in vocal anatomy across racial groups underlies any voice-memory bias. This hypothesis predicts that Black listeners will have an advantage for the Group 3 (White-sounding Black) voices. On the other hand, the Dialectal-Race Hypothesis holds that asymmetric experience with differences in features of spoken dialect underlies voice-memory bias. This hypothesis predicts that it will be the White listeners who have an advantage for the Group 3 voices. Note that structure-only models of talker identification are wholly incompatible with a memory bias for Group 3 voices among White listeners.

4.1 Stimuli

In this experiment, 15 of the original 24 voices were used: the 5 Black voices most frequently categorized as sounding Black (Group 1), the 5 White voices most frequently categorized as White (Group 2) and the 5 Black voices most frequently categorized as White (Group 3). The individual talkers that made up each group are indicated in Fig. 2a. To facilitate listeners' ability to individuate the 15 voices, each voice was assigned a unique name. To insure familiarity, these names were based on the 15 most popular bisyllabic names given to male children in the United States in 1985, the year of birth of most participants in our lab at the time (Social Security Administration, 2008).

4.2 Participants

Two new listener groups were trained to identify individuals in the three groups of voices by name. The Black listener group consisted of 11 individuals (10 females) ages 19-26 years (M = 21.1), all of whom self-identified as Black and came from entirely African-American families. The White listener group consisted of 12 individuals (9 females) ages 18-22 yrs (M = 19.5), all of whom self-identified as White and came from entirely Caucasian-American families. No listener from Experiment 1 participated in Experiment 2. One of the talkers recorded for stimuli in Experiment 1 returned as a listener for Experiment 2, but his voice was not used in any of the identification conditions.

4.3 Procedure

Listeners learned to identify by name the individual voices in each group from 5 training sentences, and were subsequently tested on their ability to identify those voices from 5 novel (untrained) test sentences. These two sets of sentences are indicated in Appendix A. Each listener participated in all three conditions (Group 1, Group 2, Group 3) with the order counterbalanced across participants. The procedure for each condition was identical, excepting the different voices to be learned. Participants were not told they would be hearing voices of different races, only that they would meet fifteen voices five at a time. Participants were offered a short break after completing each condition.

The procedure for the identification experiment is based on that of our previous work, which was effective at familiarizing vocal identity and assessing voice recognition abilities in a single experimental session (Perrachione & Wong, 2007). Participants were first familiarized with the five voices in one talker group. Listening on a pair of headphones, they heard a recording of each voice reading a sentence while the name associated with that voice was displayed on a computer screen. After hearing recordings from all five talkers, the participants practiced identifying who was who. Participants heard a recording of one of the five talkers and then indicated that talker's name by pressing the corresponding button on a button box. During practice participants always received feedback, and the computer indicated whether their response was correct, and, if not, what the correct answer should have been. After practicing identifying all five voices from one sentence, participants listened to and practiced recognizing the talkers on the remaining training sentences. After practicing all five training sentences, participants undertook a Talker Identification Test. During this test, the five untrained test sentences spoken by each of the five voices were presented twice in random order (50 trials). Listeners again indicated the name of the individual speaking by pressing the corresponding button. Participants did not receive any feedback during the Talker Identification Test. After completing the test on one talker group, participants completed the remaining conditions in turn.

Only participants' responses on the final Talker Identification Test from each condition were analyzed. Identification accuracy data were collected and analyzed with a 2×3 repeated

measures ANOVA for main effects and interactions of listener race and talker group. Accuracy was computed as the number of correct trials out of total trials presented in each condition.

4.4 Results and Discussion

Participants' accuracy on this talker-identification task is illustrated in Fig. 6a. A 2 × 3 repeated-measures ANOVA with Listener Race (Black vs. White) as a between-subjects factor, and Talker Group (Group 1 vs. Group 2 vs. Group 3) as a within-subjects factor revealed a significant Listener Race × Talker Group interaction [$F(2,20) = 7.494$, $p < 0.002$], indicating that White listeners were better at identifying Group 1 voices, and Black listeners were better at identifying Group 2 voices than vice-versa. This effect is the first empirical evidence for an own-race bias in voice identification. There was also a main effect of Talker Group [$F(2,20) = 5.914$, $p < 0.006$], owing to better overall performance on the Group 3 voices (79.7%) versus either of the other two groups (Group 1 = 71.2%, Group 2 = 69.9%). There was no effect of Listener Race ($p = 0.818$).

The Group 3 voices are the critical case for testing whether the own-race bias effect described above arises from anatomically-based or dialectally-based racial identity. Recall that the meaning of an "own-race bias" is that listeners find voices/faces of the same race as themselves easier to identify than those of another race. The main effect of Talker Group from the original ANOVA indicated that the Group 3 voices were overall easier to identify, which makes this statistic insufficiently sensitive and precludes being able to determine the underlying cause of the own-race bias based on the ANOVA alone. (Note that, had there been no main effect of Talker Group, the result would not have been obfuscated in this way.) Instead, we investigated whether participants' performance on the Group 3 voices was more similar to their performance on the Group 1 or Group 2 voices by comparing the relative strength of the participant-wise correlation coefficients across these pairs of groups. This analysis approach is based on the premise that the correlation between performance on two tasks should be stronger if they draw from similar mechanisms or abilities versus more dissimilar tasks. For the present data, if participant performance identifying individual voices in Group 3 is more closely correlated with performance on voices in Group 1, this indicates listeners behave similarly on voices that share similar vocal anatomy (i.e., are members of the same self-identified racial group), and thus the own-race bias we observed is based in asymmetric exposure to different physical (structural) features of vocal anatomy. On the other hand, if identification accuracy on Group 3 voices is more closely correlated to Group 2 voices, this indicates listeners behave similarly on voices that share similar dialectal features (i.e., are members of the same listener-categorized racial group), and thus the own-race bias is based on asymmetric exposure to these different dialectal features.

To test which of the other two groups participants' performance on the Group 3 voices more closely resembled, we correlated individual participants' accuracy between these group pairs and compared the strength of the correlations (Fisher, 1921). These results are illustrated in Fig. 6b,c. This test demonstrated that the correlation between groups sharing listener-categorized racial identity (Group 3 and Group 2; $r = 0.722$) was significantly stronger than between groups sharing self-described racial identity (Group 3 and Group 1; $r = 0.361$) [$z = 1.687$, $p < 0.05$]. These results are directly compatible with the Dialectal-Race Hypothesis, in which the own-race bias results from asymmetric exposure to culturally-acquired features of spoken dialect. Moreover, these results argue strongly for the model of person identification we described above (Fig. 1), which can account for perceptual phenomena in talker identification resulting from both vocal structure and, critically, the dynamic features of spoken language.

5. General Discussion

Individuals are perceived and identified by not only the invariant structural properties of their voice but also the dynamic features of their speech and vocal expressions. When asked to categorize voices by race, listeners are primarily sensitive to the dynamic, socially-acquired features of a talker's speech as opposed to features attributable to vocal structure. As such, self-descriptions of racial identity, which are based primarily on the physical (visual) features socially associated with race, are often disjoint from listeners' perceptions of race based on auditory information alone. When identifying individual voices of different racial groups, listeners gain an advantage from enhanced experience with features of spoken dialect rather than differences in vocal anatomy across those groups. These results argue strongly for the Dialectal-Race Hypothesis we outlined, as well as specifically against the Anatomical-Race Hypothesis, and, therefore, against structure-only models of person perception that do not account for cultural asymmetries in relevant socially-acquired distinguishing features. This is further supported by the results of our acoustic analyses, which demonstrated that socially-acquired phonetic features of spoken dialect by themselves are consistent with listeners' perceptions of race from voice. This does not mean, however, that there is no place for structural analysis in talker identification. In fact, structural analysis is a critical first step in voice recognition, both generally for conspecifics and specifically for individuals. However, structure-only models are insufficient in accounting for a wide variety of cognitive phenomena in talker identification, such as memory biases resulting from asymmetric exposure to phonetic variation.

The own-race bias (or, more accurately, the own-dialect bias) in talker identification is the result of asymmetric cultural experiences with speech of different talker groups. In asymmetric cultural experiences, the quantity or quality of some experiences exceeds that of others. Since individuals predominately associate with others with whom they share (or, for children, will come to share) a linguistic background, their auditory system becomes primarily tuned to the meaningful variation within that language or dialect. In different dialects, this variation may encompass different regions of perceptual space – as we show here for dimensions of consonant voicing and vowel fronting – and limited experience with such variability can reduce its informativeness for other auditory tasks, such as talker identification. The effects of such asymmetric experiences on sensitivity and attention can be seen even very early in auditory development (Kinzler, Dupoux, & Spelke, 2007).

Unlike face perception, which has been studied primarily using the static features of faces present in photographs, ecological person identification from voice necessarily relies on the dynamic cues present in a rapidly varying acoustic signal. Further research is needed to understand what role dynamic cues may play in face perception. Indeed, visual dynamic cues alone are a sufficient indicator of personhood, as in the perception of biological motion from point-light displays (Neri, Morrone, & Burr, 1998). Person identity, as a social construct independent of sensory modality, is highly adaptive to cultural experiences, and our findings challenge the existing notion that structural features are necessarily the primary basis for person identification in all domains. Our revised model (Fig. 1) illustrates how mechanisms mediating perception of culturally-acquired features of expression also critically contribute to person perception abilities. In addition to accounting for how differential exposure to culturally-acquired dynamic features of speech has a significant influence on person perception abilities in the auditory domain, this model also provides an extended framework for testing how analogous dynamic features may affect person identification in other sensory domains.

In addition to a broadened scientific understanding of the mechanisms of person perception, these results also have practical implications for educational, medical, forensic, and social policies and practices. Racial bias and discrimination create substantial injustice across social

domains, including employment practices, judicial proceedings, and the allotment of civil resources. Our results show that the perception of race is not always tied to individuals' physical characteristics, but is often an attribute deduced from socially-acquired expressive behaviors, such as features of spoken dialect. Previous work has shown that specific phonetic features covary by race and dialect, and our current results suggest that listeners' perceptions of race from voice are indeed consistent with these dialectal differences. These findings are especially significant given the differences in behaviors and attitudes associated with biological versus social conceptualizations of race (Williams and Eberhardt, 2008). Understanding how physical and expressive features interact to give rise to the perception of race will greatly enhance our ability to become familiar with the social and cultural distinctions that may, otherwise unrecognized, give rise to undue bias and prejudice.

Acknowledgments

We thank Geshri Gunasekera, Johnston Chen, Louisa Ha, Peter Hsieh, and Tasha Dees for their assistance conducting this research. We are grateful to Stefanie Shattuck-Hufnagel and, especially, the editor and anonymous reviewers for their invaluable comments and advice on the manuscript. Portions of this work were presented at the 155th meeting of the Acoustical Society of America, 2008 (Paris, France). This work was supported by the National Institutes of Health (USA) grants R01DC008333, R03HD051827, R21DC009652 and R21DC007468, and National Science Foundation grants BCS-0719666 to P.W. and BCS-0720312 and BCS-0722326 to J.C.

Appendix A: Stimuli Sentences

These sentences (IEEE, 1969) were read by all our talkers. Listeners in Experiment 1 categorized the voices from all 10 of these sentences. Listeners in Experiment 2 heard the first 5 sentences during training, and the remaining 5 sentences during the test.

Training Sentences

1. The boy was there when the sun rose.
2. A rod is used to catch pink salmon.
3. The source of the huge river is the clear spring.
4. Kick the ball straight and follow through.
5. Help the woman get back to her feet.

Testing Sentences

1. A pot of tea helps to pass the evening.
2. Smoky fires lack flame and heat.
3. The soft cushion broke the man's fall.
4. The salt breeze came across from the sea.
5. The girl at the booth sold fifty bonds.

Appendix B: Talker Characteristics

To control for non-race-related differences between the two talker groups in Experiment 1, these 24 voices were matched for height, weight, and age. The two groups did not differ significantly in any of these measures (see main text)

Black Talkers				White Talkers			
Subject	**Height (inches)**	**Weight (pounds)**	**Age (years)**	**Subject**	**Height (inches)**	**Weight (pounds)**	**Age (years)**
6-166	71	150	20	5-105	70	140	24
6-181	67	137	18	7-055	69	125	18
7-005	72	155	22	7-057	73	150	20
7-012	74	203	21	7-061	74	200	22
7-016	69	195	20	7-056	70	200	19
7-019	67	185	21	7-060	71	170	19
7-037	71	190	32	5-103	70	175	27
7-041	66	160	20	5-072	66	170	20
7-042	74	260	17	7-058	73	145	18
7-045	67	154	19	5-102	66	135	20
7-049	72	210	18	7-051	72	145	19
7-050	75	185	18	7-059	76	185	18
Mean =	**70.42**	**182.00**	**20.50**	**Mean =**	**70.83**	**161.67**	**20.33**
Std. Dev. =	**(3.15)**	**(33.85)**	**(3.92)**	**Std. Dev. =**	**(3.01)**	**(25.26)**	**(2.74)**

References

Allen JS, Miller JL. Listener sensitivity to individual talker differences in voice-onset-time. Journal of the Acoustical Society of America 2004;115:3171–3183. [PubMed: 15237841]

Belin P, Fecteau S, Bédard C. Thinking the voice: Neural correlates of voice perception. Trends in Cognitive Science 2004;3:129–135.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature 2000;403:309–312. [PubMed: 10659849]

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer. 2008. (Version 5.0.32) [Computer program]. http://www.praat.org/

Bruce V, Young A. Understanding face recognition. British Journal of Psychology 1986;77:305–327. [PubMed: 3756376]

Bruckert L, Liénard JS, Lacroix A, Kreutzer M, Leboucher G. Women use voice parameters to assess men's characteristics. Proceedings of the Royal Society of London B -Biological Sciences 2006;273:83–89.

Campanella S, Belin P. Integrating face and voice in person perception. Trends in Cognitive Sciences 2007;11:535–543. [PubMed: 17997124]

Chiao, JY.; Ambady, N. Cultural neuroscience: Parsing universality and diversity across levels of analysis. In: Kitayama, S.; Cohen, D., editors. Handbook of Cultural Psychology. Guilford Press; New York: 2007. p. 237-254.

Chiao JY, Iidaka T, Gordon HL, Nogawa J, Bar M, Aminoff E, Sadato N, Ambady N. Cultural specificity in amygdala response to fear faces. Journal of Cognitive Neuroscience 2008;20:2167–2174. [PubMed: 18457504]

Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. Journal of Phonetics 2004;32:111–140.

Clopper CG, Pisoni DB, de Jong K. Acoustic characteristics of the vowel systems of six regional varieties of American English. Journal of the Acoustical Society of America 2005;118:1661–1676. [PubMed: 16240825]

Craig HK, Thompson CA, Washington JA, Potter SL. Phonological features of child African American English. Journal of Speech, Language, and Hearing Research 2003;46:623–635.

Downing PE. A cortical area selective for visual processing of the human body. Science 2001;293:2470–2473. [PubMed: 11577239]

Evans BG, Iverson P. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British 99English sentences. Journal of the Acoustical Society of America 2004;115:352–361. [PubMed: 14759027]

Evans S, Neave N, Wakelin D. Relationship between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. Biological Psychology 2006;72:160–163. [PubMed: 16280195]

Fisher RA. On the probable error of a coefficient of correlation deduced from a small sample. Metron 1921;1:1–32.

Golby AJ, Gabrieli JDE, Chiao JY, Eberhardt JL. Differential responses in the fusiform region to same-race and other-race faces. Nature Neuroscience 2001;4:845–850.

Green, LJ. African-American English: A linguistic introduction. New York: Cambridge University Press; 2002.

Hayward WG, Rhodes G, Schwaninger A. An own-race advantage for components as well as configurations in face recognition. Cognition 2008;106:1017–1027. [PubMed: 17524388]

Institute of Electrical and Electronics Engineers. IEEE recomended practices for speech quality measurements. IEEE Transactions on Audio and Electroacoustics 1969;17:225–246.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience 1997;17:4302–4311. [PubMed: 9151747]

Kinzler KD, Dupoux E, Spelke ES. The native language of social cognition. Proceedings of the National Academies of Science 2007;104:12577–12580.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 1992;255:606–608. [PubMed: 1736364]

Lass NJ, Brown WS. Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. Journal of the Acoustical Society of America 1978;63:1218–1220. [PubMed: 649880]

Lass NJ, Tecca JE, Mancuso RA, Black WI. The effect of phonetic complexity on speaker race and sex identifications. Journal of Phonetics 1979;7:105–118.

Meissner CA, Brigham JC. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. Psychology, Public Policy, and Law 2001;7:3–35.

Neri P, Morrone MC, Burr DC. Seeing biological motion. Nature 1998;395:894–896. [PubMed: 9804421]

Pascalis O, Scott LS, Kelly DJ, Shannon RW, Nicholson E, Coleman M, Nelson CA. Plasticity of face processing in infancy. Proceedings of the National Academies of Science 2005;102:5297–5300.

Perrachione TK, Wong PCM. Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. Neuropsychologia 2007;45:1899–1910. [PubMed: 17258240]

Pollock KE, Meredith LH. Phonetic transcription of African American Vernacular English. Communication Disorders Quarterly 2001;23:47–53.

Purnell T, Idsardi W, Baugh J. Perceptual and phonetic experiments on American English dialect identification. Journal of Language and Social Psychology 1999;18:10–30.

Remez RE, Fellowes JM, Rubin PE. Talker identification based on phonetic information. Journal of Experimental Psychology: Human Perception and Performance 1997;23:651–666. [PubMed: 9180039]

Social Security Administration. Popular baby names. 2008. http://www.ssa.gov/OACT/babynames/

Thomas ER, Reaser J. Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. Journal of Sociolinguistics 2004;8:54–87.

Williams MJ, Eberhardt JL. Biological conceptions of race and the motivation to cross racial boundaries. Journal of Personality and Social Psychology 2008;94:1033–1047. [PubMed: 18505316]

Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nature Neuroscience 2007;10:420–422.

Wong PCM, Parson LM, Martinez M, Diehl RL. The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. Journal of Neuroscience 2004;24:9153–9160. [PubMed: 15483134]

Xue SA, Hao GJP, Mayo R. Volumetric measurements of vocal tracts for male speakers from different races. Clinical Linguistics and Phonetics 2006;20:691–702. [PubMed: 17342877]

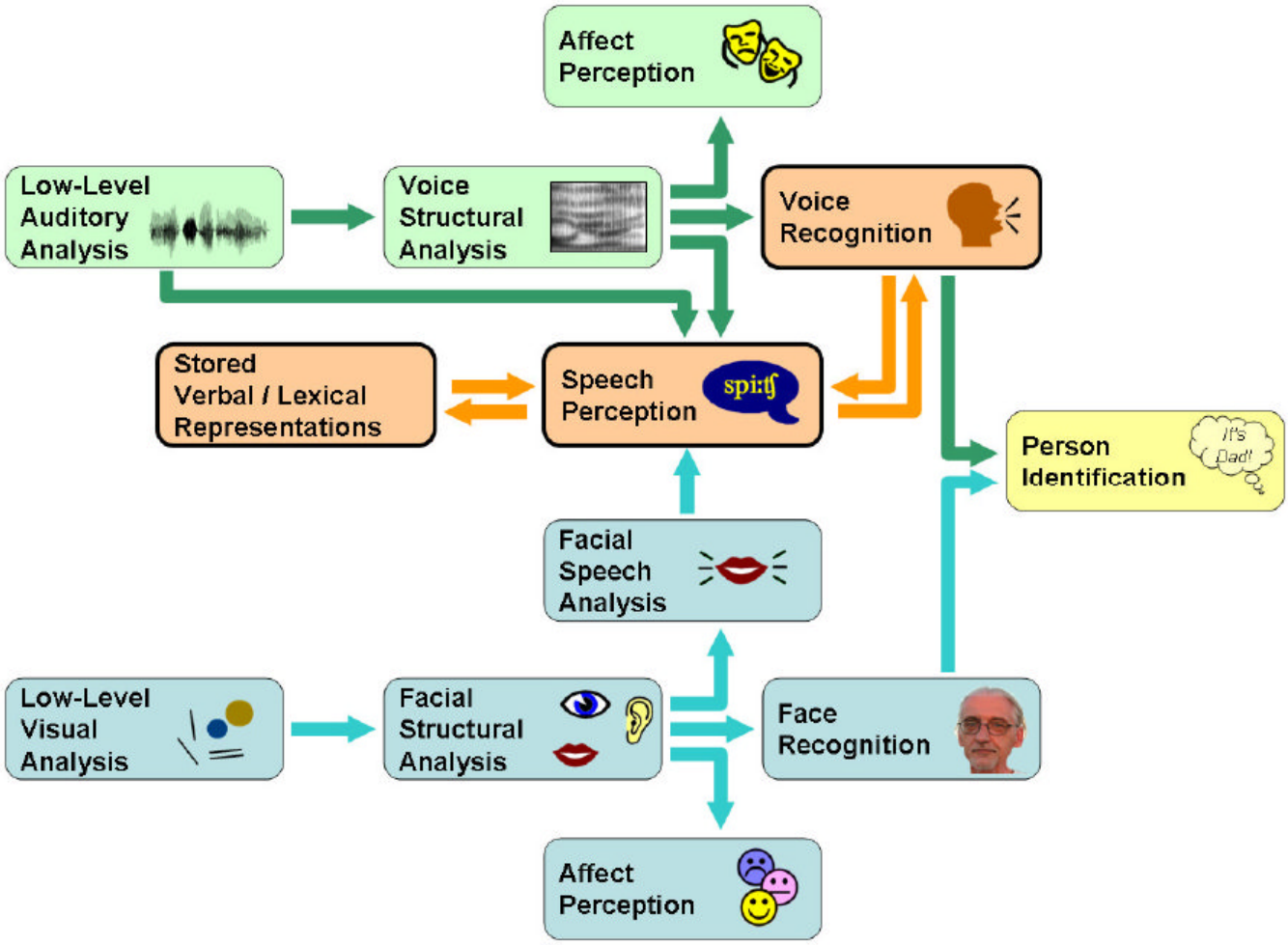

Figure 1.

An integrated model of person perception from face and/or voice. Boxes represent perceptual, computational, or mnemonic modules. Arrows represent major directional pathways of shared information. Blue boxes and arrows denote the face-perception system. Green boxes and arrows denote the voice-perception system. Orange arrows indicate important routes of shared information in the voice-perception system lacking analogous connections in the face-perception system. Boxes with bold borders are those modules that rely on shared connections unique to the voice-perception system. The "Speech Perception" component should be broadly construed to represent all idiosyncratic features of the talker independent from the structure of his or her vocal tract. (This model illustrates only empirically supported pathways relevant to person identification).

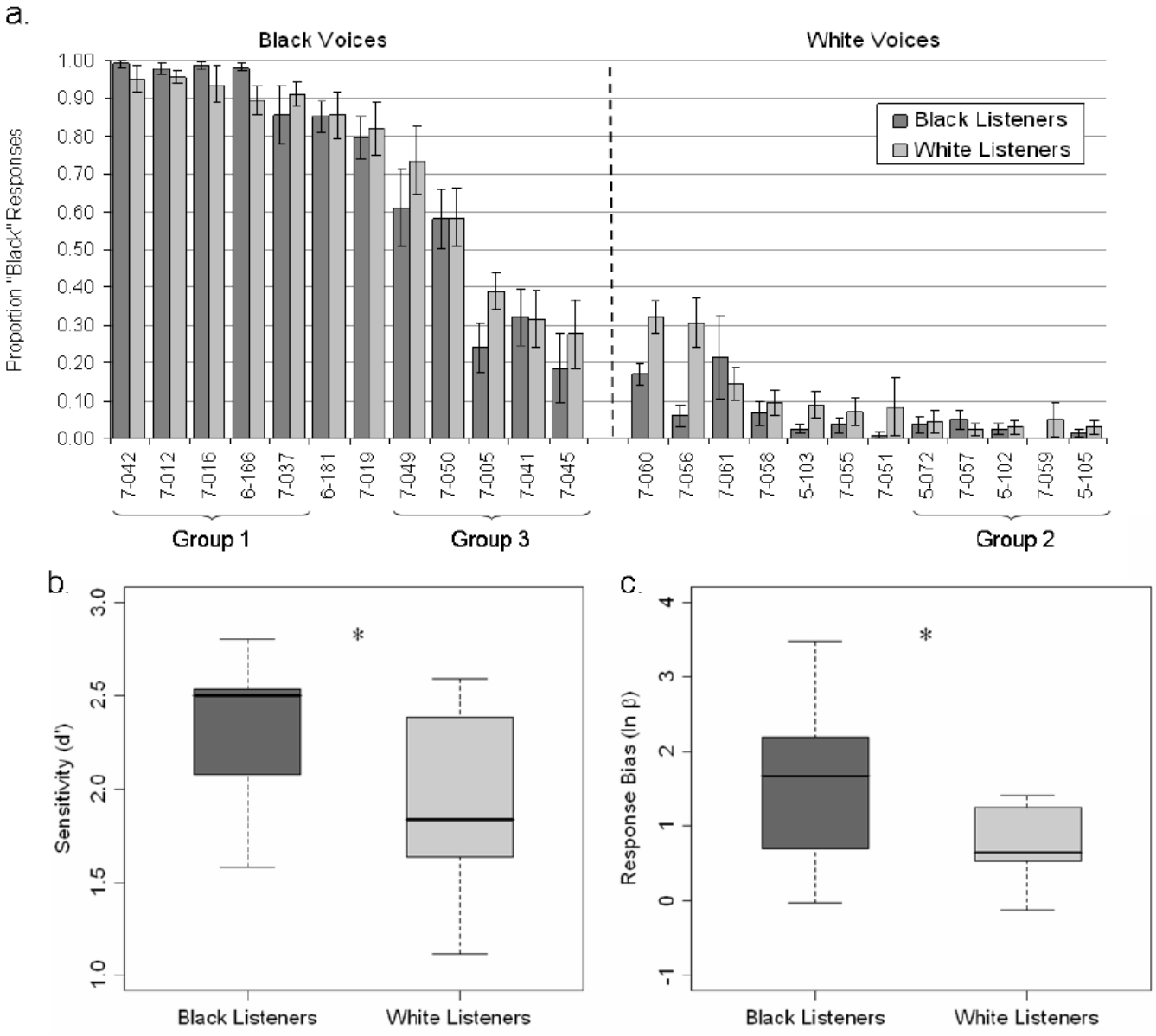

Figure 2.

Racial categorization of voices by listener group. **(A)** Proportion of "Black" responses for each voice by Black (dark bars) and White (light bars) listeners. Black voices are shown on the left, White voices on the right. Error bars represent standard error of the mean. Voices are ordered by the overall frequency of categorization as "Black." Brackets below the abscissa indicate voices used in the three talker groups in Experiment 2. **(B)** The Black listener group was significantly more sensitive to the presence of Black racial information in the categorization experiment ($p < 0.05$). A dark horizontal line indicates the mean, the filled rectangles encompass the interquartile range, and high-low bars indicate maximum and minimum points. **(C)** Black listeners were significantly more conservative than White listeners in their categorization of voices as sounding Black ($p < 0.035$).

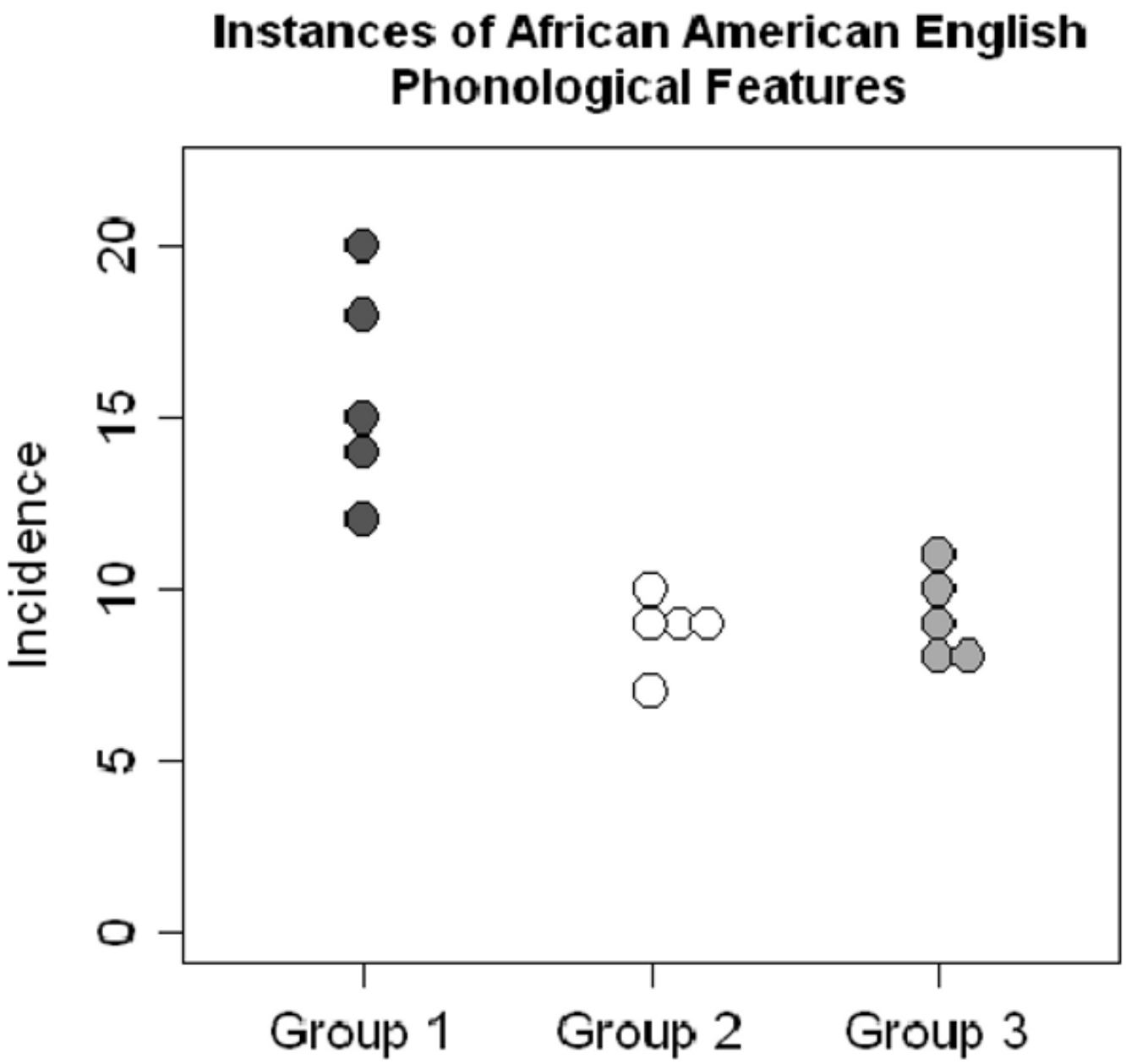

Figure 3.

The three talker groups were distinguished by the incidence of common phonological features of African-American English dialects. The distribution of these features was significantly more predictive of listener-perceived race (shared by Groups 2 and 3, $p = 0.0003$) than talkers' self-identified race (shared by Groups 1 and 2, $p = 0.767$), consistent with the idea that race and dialect are incongruous among the three groups. Additionally, the frequency of phonological feature types varied more between Group 1 and the other two groups than between Groups 2 and 3. Circles represent number of instances of these phonological features in the stimuli from Experiment 1 for each individual in a group. Shading differs by group. Points of equal value spread along the abscissa to avoid overlap.

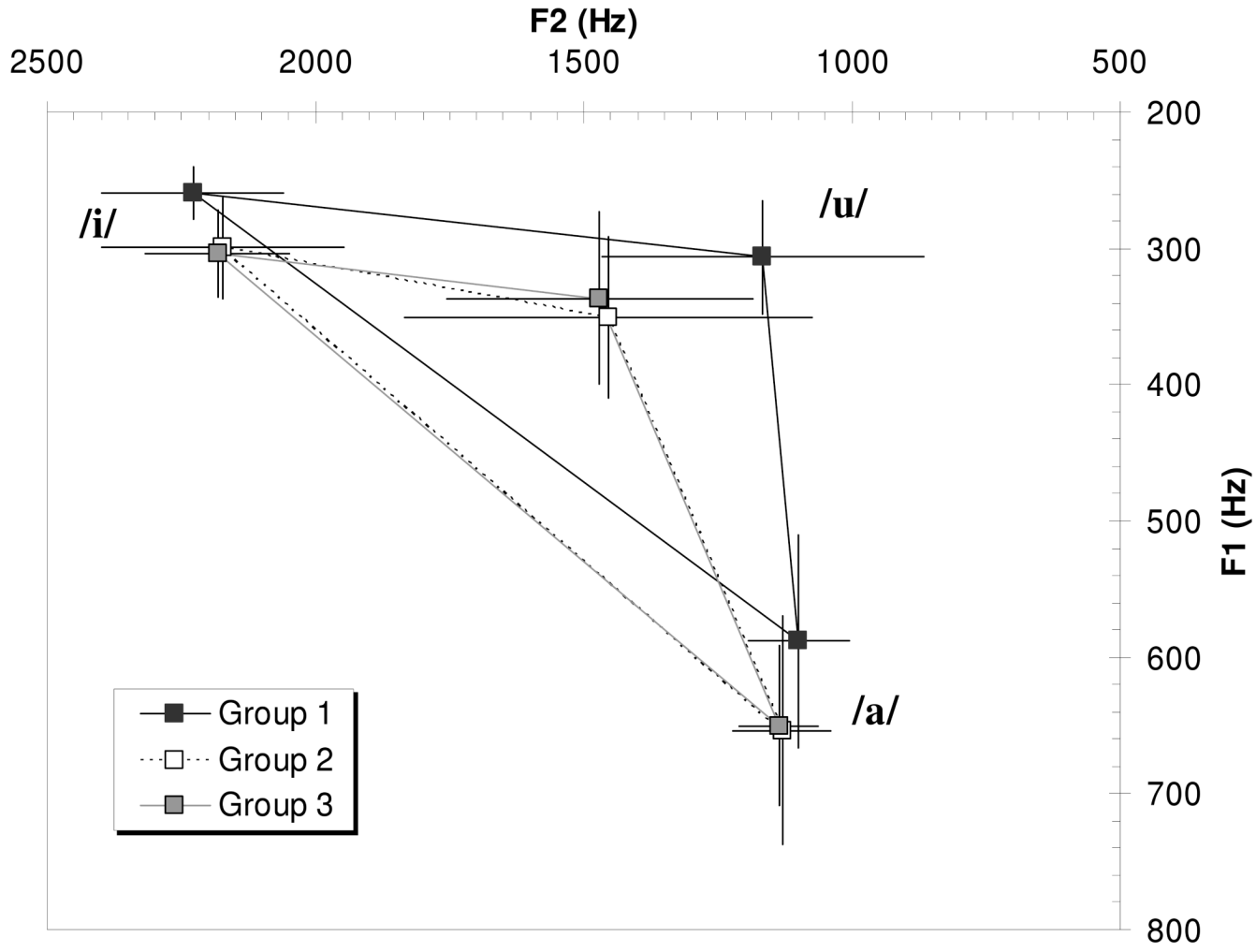

Figure 4.
Vowel spaces of the three talker groups based on the cardinal vowels /a/, /i/, and /u/ (as in "pot", "heat", and "booth"). Points indicate the mean locus of each vowel across all tokens in a group; error bars indicate one standard deviation of the mean. In addition to a consistently higher F1, Groups 2 and 3 exhibited significant fronting of the vowel /u/, characteristic of White dialects of American English, compared to Group 1. The vowel space of Group 3 talkers was virtually identical to Group 2 talkers, consistent with the pattern of listener-categorized racial identity (Fig. 2).

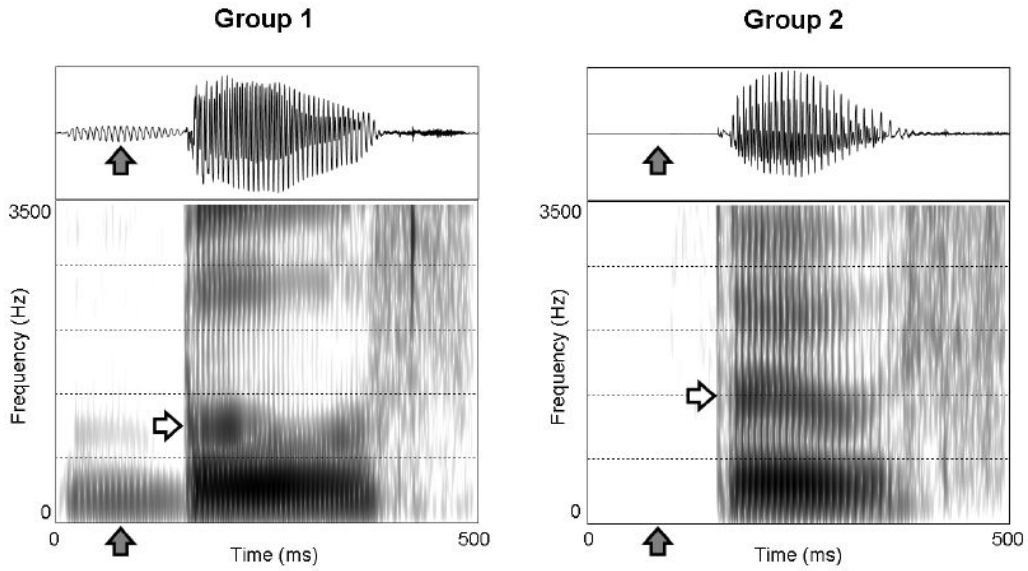

Figure 5.

Productions of the word “booth” (/buθ/) by individual talkers in Group 1 and Group 2 which exemplify two of the distinctive dialectal phonetic features examined. As indicated by the filled arrows, both the waveform (top) and spectrogram (bottom) of the Group 1 talker prominently illustrate the prevoicing (negative voice-onset time) previously associated with African-American dialects of American English. This contrasts with the Group 2 talker, where no prevoicing is evident. The open arrow indicates the location of the second formant (F2) on the spectrogram. As shown in Fig. 4, Group 2 talkers had a significantly higher F2 in the vowel /u/ than Group 1 talkers. This distinction is also evident in the two example talkers' productions shown here.

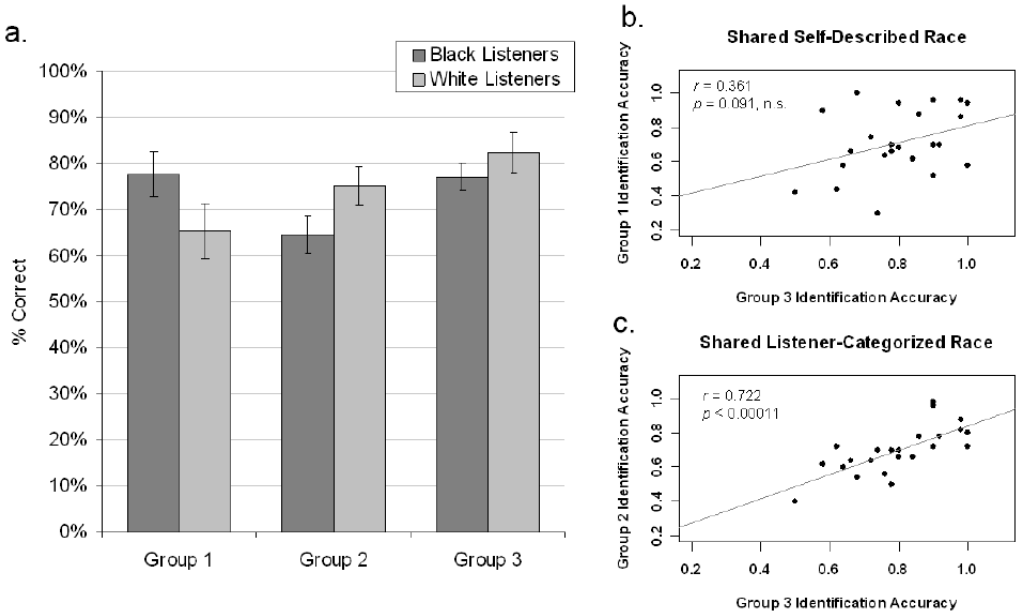

Figure 6.

Talker identification test accuracy. **(A)** Both Black and White listeners displayed an advantage for identifying Group 1 and Group 2 voices, respectively ($p < 0.002$) – the hallmark of an own-race bias effect. Group 3 voices were more accurately identified by both groups ($p < 0.006$), but further analysis revealed performance on these voices was more similar to Group 2 voices than Group 1 voices. Error bars indicate standard error of the mean. **(B)** The correlation between participants' pooled performance on identifying Group 3 voices and Group 1 voices (which shared self-described racial identity) was not reliably significant [Pearson's $r = 0.361$, $p = 0.091$]. **(C)** The correlation between identification performance on Group 3 voices and Group 2 voices (shared listener-categorized race) was significant [Pearson's $r = 0.722$, $p < 0.0001$]. The correlation between groups sharing listener-categorized race (shared features of spoken dialect, Groups 2 & 3) was reliably stronger ($p < 0.05$) as determined by computing the difference of the Fisher-z transformed coefficients.