Influence of Suprasegmental Features on Perceived Ethnicity of American Politicians

Nicole R. Holliday and Zachary S. Jaggers

Department of Linguistics, New York University
nicole.holliday@nyu.edu; zackjaggers@nyu.edu

ABSTRACT

How accurate are listeners at identifying the ethnicities of political figures from one-word samples? Do suprasegmental variables provide a basis for these judgments? Tokens of six lexical items were extracted from speeches by seven male political figures of different stated ethnic identities. In a Mechanical Turk experiment, 94 listeners heard each token twice, then responded to the multiple-choice question: “What is the ethnicity of this speaker?”. While listeners overall performed with an accuracy rate below chance, certain speakers and tokens were more consistently identified than others, both accurately and inaccurately. Analysis indicates suprasegmental features including Intensity Ratio, Harmonics to Noise Ratio, Jitter Average. Pitch Peak Ratio, and Syllable Duration Ratio, contribute to listener judgments of ethnicity.

Keywords: sociophonetics, variation, ethnicity, speech perception

1. INTRODUCTION

Perception of dialects as ethnically linked has been shown to have important social and economic consequences for speakers across the United States [1, 4]. In fact, earlier research found that a voice in an African American or Chicano guise was substantially less likely to receive an appointment to view an apartment than a voice in a Standard American English guise [1, 5]. Despite this fact, few studies have examined the specific acoustic properties that contribute to a voice’s identification with a particular ethnic group, and even fewer have done so with naturalistic speech.

In a matched-guise listening task, Purnell, Idsardi, and Baugh [5] found that listeners were over 70% accurate at differentiating between the following English dialects: Chicano English, African American English, Standard American English, based only on hearing a stimulus of the token *hello*, as produced by Baugh in various guises. This seems to indicate that listeners make fairly fast and accurate judgments about a speaker’s ethnicity, but this result has not been tested for speakers who were not explicitly instructed to perform a particular guise. The current study examines whether naïve listeners can make accurate judgments about speaker ethnicity based on one-word samples extracted from political speeches. It also tests the influence of several suprasegmental features on ethnic identification.

2. METHODOLOGY

This study elicits listener judgments of speaker ethnicity based on bisyllabic one-word samples extracted from naturalistic speech, and then explores the contribution of phonetic properties to these judgments. Utterances of the same six lexical items were extracted from televised speeches by seven male political figures primarily from cities in the northeastern U.S. The speech of political figures is particularly useful for this type of experiment because it is easy to extract lexical items from readily available high quality recordings, and also possible to control for formality and type of speech situation in which the tokens were uttered (in this case, all speeches to mainstream audiences or to the U.S. Congress). The seven politicians used in this sample are of three different stated ethnic identities: three black (Cory Booker, Deval Patrick, David Paterson) two Latino (Bob Menendez, Marco Rubio), and two white non-Latino (Andrew Cuomo and Bill de Blasio).

Tokens of the following six high-frequency items were extracted from their public speeches: *city, many, people, women, issues, thank you*. These stimuli were chosen with the purpose of capturing how suprasegmental features contribute to listener identification of speaker ethnicity and so, to the extent possible, these items avoid some potential sites of segmental variation associated with AAE (e.g., coronal stop deletion, [aɪ] monopthongization), or urban Northeastern US dialects (e.g., non-rhoticity, COT/CAUGHT contrast). The brevity of the speech samples presented also aided in minimizing potential influence of segmental features.

The 42 tokens (six tokens from each of the seven speakers) were compiled and presented in an individually randomized order to participants via a Mechanical Turk experiment. Listeners (N=94) heard each token twice, with a two-second pause in between repetitions. They were then given seven seconds to respond to the forced multiple-choice question “What is the ethnicity of this speaker?”. Partially following the methodology of Purnell et al. [5], the choices provided were “black”, “Latino” and “white” in order to constrain the realm of possible choices and elicit quick judgments.

Listeners were asked at the midpoint of the experiment to type in the brand of headphones they were using in order to further confirm that they were indeed still listening and participating. After the listening task, participants responded to a short demographic questionnaire including questions about their age, region, gender and ethnicity in order to test for potential effects of listener demographics on perceptual judgments. Finally, at the conclusion of the experiment, listeners were also asked to rate the difficulty of the task and to note if they had recognized any of the voices in the sample. Participants found the task moderately difficult overall (M=4.87 out of 10), and none of the 94 participants accurately identified any of the voices, eliminating potential effects of individual speaker recognition on the results.

3. ACOUSTIC ANALYSIS

Earlier literature has described prosodic and voice quality differences between voices likely to be associated with different ethnic groups (see [6] for a thorough discussion). Scholars have tended to focus on potential suprasegmental differences between black and white Americans, though there is little literature that specifically examines the acoustic properties that lead to these judgments. Purnell et al. [5] did identify some suprasegmental phonetic features as significant correlates with ethnolect identification. Specifically, they observed that Pitch Peak Ratio and Harmonics to Noise Ratio were highest for black-identified voices. The current study further investigates which suprasegemental features differ between voices identified with a particular ethnicity.

The utterance tokens extracted from the political speeches were analysed along a number of such variables. First, all tokens were segmented in Praat software and divided into syllables. The SYLLABLE DURATION RATIO (or SDR) was calculated, dividing the duration of the first syllable by that of the second, as well as the INTENSITY RATIO, measuring the difference in intensity (a function of both amplitude and frequency) of the two syllable nuclei. The rate of change in pitch, the PITCH PEAK RATIO (or PPR), was calculated by dividing the difference between the F0 max and F0 min by the duration of time between their points of occurrence.

For features of voice quality, the following were averaged across the two syllable nuclei: SHIMMER (“local” version in Praat), the average of local variation in amplitude between vocal fold vibrations; JITTER (“rap” in Praat), the average of local variation in F0 between vocal fold vibrations; HARMONICS TO NOISE RATIO (or HNR), the ratio of periodic components to aperiodic components; and INTENSITY AVERAGE, average of amplitude as a function of frequency.

4. RESULTS

4.1. Listener Accuracy and Demographics

In general, listeners were not accurate at identifying the ethnicity of the speakers, with most performing at a rate near chance (M=28.57%) between the three options (black, Latino, white). Listener demographics of Age, Region, Gender, and Ethnicity were included in a step-up/step-down logistic regression analysis in Rbrul on the response token set (N=3948) to see if any demographic groups performed with higher accuracy than others. Though one advantage of Mechanical Turk is a potentially diverse group of respondents, it is difficult to control a priori for all demographic characteristics. Participants were asked to self-identify with one of six age groups, which were as follows: 18-24, 25-34, 35-44, 45-54, 55-64, 65+. The age groups were not evenly distributed, and the sample tended toward those in the 25-44 range (62% of participants). Overall, Gender was well balanced (49% female, 51% male) in the respondent pool. Results for Region were spread across the US with listeners reporting relatively even numbers from the given categories of Midwest, Mid-Atlantic, New England, West Coast, Southeast, and Southwest. Unfortunately listener ethnicity could not be reliably tested, as the sample included a disproportionate number of white respondents (81%), which is comparable to general Mechanical Turk demographics [2]. Ultimately, none of the demographic factors tested was found to significantly affect accuracy or frequency of rating tokens as black vs. other. Additionally, Individual Listener was analyzed as an independent variable and also found to have no effect, suggesting no individuals were significantly more accurate or more biased toward one option.

4.2. Accuracy (and Inaccuracy) by Speaker

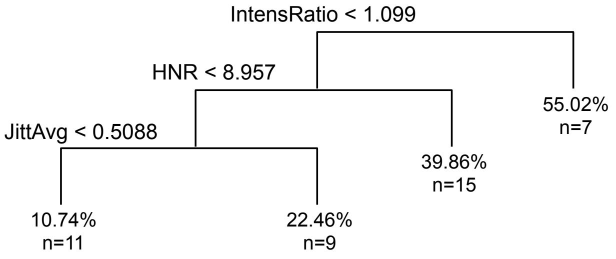
Averaging the responses elicited by the tokens of the seven speakers, the results of four were near chance (M=31.26%), suggesting listeners did not strongly identify their speech with any ethnicity. Despite this general trend, the results of three of the politicians in the sample were notably more consistent than the overall results. Cory Booker was reliably and accurately identified as black (72%). Bill de Blasio was reliably and accurately identified as white (57%), and infrequently as black (20%, a rate less than chance). Andrew Cuomo, however, was reliably *mis*identified as black (68%) and very infrequently identified (accurately) as white (15%). Both the accurately and inaccurately black-identified speakers were also the most consistently identified, paralleling Purnell et al.'s [5] discussion that listeners may consider blackness as more perceptibly distinct than other ethnicities. Furthermore, the fact that no speaker was accurately or inaccurately identified as Latino with any consistency further supports the notion of binary perception of ethnicity. The next section attempts to identify what suprasegmental phonetic correlates there may be with this percept.

4.3. Effects of Acoustic Variables

To test how acoustic factors influenced listeners' responses to the question of speaker ethnicity, a multivariate stepwise logistic regression analysis was performed using the Rbrul package in R statistical software. Since the primary aim is to see what features listeners associate with black speakers, this analysis tested whether or not a token was identified black as the binary dependent variable with the acoustic factors defined above and Speaker as independent variables. A step-up/step-down analysis performed per response selected the following significant factors (p<.05): SDR, Intensity Ratio, PPR, Jitter, HNR, and Intensity Average (i.e., all the acoustic variables except Shimmer). Speaker was not chosen as significant, suggesting that acoustic factors, rather than actual speaker, better predict responses.

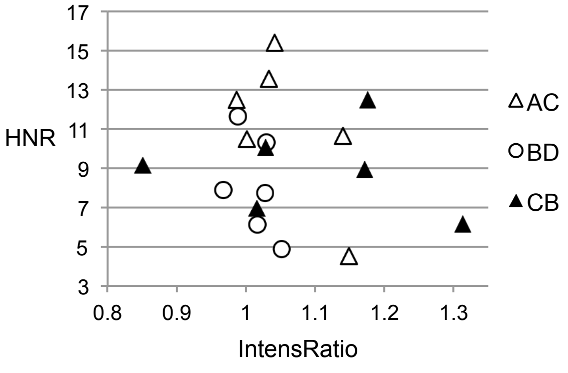
These results confirm that suprasegmental features play a role in judgments of speaker ethnicity. However, the fact that several factors were found to have a significant effect leaves the open-ended question of whether any particular variables may play a stronger role than others. To address this, a Correlation and Regression Tree (CART) analysis was performed using the Rpart package in R. This analysis tested how well different acoustic variables did at categorizing the 42 utterance tokens into exclusive groups based on their frequency of being identified as black. The results in Figure 1 show that the most predictive variable was higher Intensity Ratio. The second and third most predictive factors were higher HNR and higher Jitter Average. Note that this model works at the group level but is not predictable at the individual token level.

Figure 1: CART diagram. Node heads indicate group criteria. Node ends represent number of tokens within each group and average black identification frequency.



To see how the predictions of this test fare at the level of individual utterances, tokens of the three most consistently (mis)identified speakers were plotted along Intensity Ratio and HNR, shown in Figure 2. By the results of the CART analysis, tokens falling in the upper right quadrant of this plot would be predicted as most frequently identified as black. First comparing Andrew Cuomo (AC) and Bill de Blasio (BD), AC's tokens indeed trend more toward the upper right than BD's, in line with AC's frequent misidentification as black. Cory Booker's (CB) tokens, however, appear more widespread along these two features. Booker has tokens with high Intensity Ratio and HNR, but overall his tokens do not cluster more toward the upper right than those of AC and BD, as might have been expected. Tokens of CB's which did fall toward the lower left of this plot were still consistently identified as black, suggesting that these features alone do not fully predict black identification.

Figure 2: Plot of Andrew Cuomo, Bill de Blasio, and Cory Booker's tokens by the two strongest acoustic predictors of black identification.



5. DISCUSSION

These results demonstrate that the mechanism that listeners use to make judgments of ethnicity based on one-word tokens is complex and likely operates on a token-by-token and speaker-by-speaker basis. In particular, the results obtained for the tokens from Cory Booker, Andrew Cuomo, and Bill De Blasio show that though particular acoustic properties seem to be important in determining whether a token is identified as black, the relationship between those properties and likelihood of a token being identified as black is not linear. The fact that some of the most reliably identified tokens from Cory Booker's speeches scored low on the most predictive properties (Intensity Ratio and HNR) indicates that listeners may be using a combination of features in making their judgments of ethnicity, as suggested by the suite of features deemed significant predictors in the Rbrul analysis.

Additionally, this study is another advance toward understanding the properties that listeners attune to in making judgments of speaker ethnicity. Purnell et al. [5] also found that first syllable duration, HNR, and location of peak F0 appear to differ between the guises in their study and thus may be useful for listeners in dialect identification. Our results further support their findings about syllable duration and HNR, and furthermore indicate that these properties seem to be salient for listeners both when they hear a voice in guise and when they are asked to judge tokens from naturalistic speech.

The result that these properties affect naturalistic speech is also particularly important for the type of naturalistic speech used in this sample. The speech of politicians has been shown to be an important element of their constructions of public personae [3], which may also have an effect on their political outcomes. Systematically identifying the salient features that characterize ethnolects for different listeners may have special consequences for these political figures, because they are often racialized explicitly and implicitly in public discourse. Knowledge of suprasegmental features that are ethnically linked could be a powerful tool for politicians, linguists, and laypeople to better understand how micro-level linguistic judgments could affect political outcomes.

6. CONCLUSION

The results of this experiment indicate that speakers generally do not perform better than chance when asked to identify the ethnicity of a speaker based on a bisyllabic one-word sample. However, it is the case that two of the seven speakers in this sample, Bill de Blasio (white) and Cory Booker (black) were accurately and reliably identified by listeners at a rate substantially higher than chance. This result seems to indicate that some speakers are much more ethnically identifiable than others, but more research is needed to determine how listeners identify these speakers. Andrew Cuomo (white) being consistently misidentified as black motivates further analysis of his speech style in relation to his northeastern U.S. Italian-American identity and examination of his use of other potential AAE-associated features. However, these results suggest that suprasegmental properties influence listener judgments of speaker ethnicity, whether these result in accurate judgments or not.

Acoustic analyses of several suprasegmental features of the tokens in this sample reveal some trends that appear to influence the likelihood of a particular token's identification as having been uttered by a black speaker. An Rbrul analysis found Intensity Ratio, Intensity Average, Syllable Duration Ratio, Pitch Peak Ratio, and Harmonics to Noise Ratio as significant factors. When tokens were tested for likelihood of being identified as uttered by a black speaker, a Correlation and Regression Tree Analysis indicated that these properties do not appear to be weighted equally in listener judgments. In line with previous research, tokens with a higher Intensity Ratio, Harmonics to Noise ratio, or Jitter Average seem to be more likely to be identified as having been uttered by a black speaker. Additional work is needed to further explore the contributions of these acoustic properties to listener perceptions of ethnicity in a speech sample.

7. REFERENCES

[1] Baugh, J. 2003. Linguistic profiling. *Black linguistics: Language, society, and politics in Africa and the Americas*. London: Routledge, 155–168.

[2] Berinsky, A. J., Huber, G.A., Lenz, G.S. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis* 20.3, 351–368.

[3] Hall-Lew, L., Coppock, E., Starr, R.L. Indexing political persuasion: Variation in the Iraq vowels. *American Speech* 85.1, 91–102.

[4] Lippi-Green, R. 1997. *English with an accent: Language, ideology, and discrimination in the United States*. London: Routledge.

[5] 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.

[6] Thomas, E. R. Forthcoming. Prosodic Features of African American English. In: Lanehart, S.L., Green, L., Bloomquist, J. (eds), *The Oxford Handbook of*

African American Language. Oxford: Oxford University Press.