DIALECT IDENTIFICATION FROM PROSODIC CUES

Christina G. Foreman
University of California, Los Angeles, USA

ABSTRACT
It is a well-known fact that dialects differ in prosodic features as well as segmental features. Though few studies have focused on prosody, existing research provides evidence of distinct differences in the intonation pattern between African-American English (AAE) and Mainstream American English (MAE). This study examines how speakers use such prosodic cues to differentiate AAE from MAE. It was hypothesized that dialect identification is easiest in cases where the intonation pattern of an utterance is dialect-specific and does not vary across speakers within a dialect group. It was also hypothesized that the degree of exposure to AAE is positively correlated with the accuracy rate of dialect identification, but exposure to AAE had a measurable effect on performance in terms of both accuracy and confidence in identification.

1. INTRODUCTION

1.1. Introduction and Background
Dialect differences that exist among ethnic groups can be characterized by distinctions in lexicon, phonology, morphology and prosody. Looking specifically at African-American English (AAE) and Mainstream American English (MAE) dialects, even the untrained listener is able to detect some distinction between these two dialects in their intonation pattern. Such apparent distinctions gave rise to a body of literature that focused on providing descriptions of AAE production. Initial studies, such as Tarone (1972) provided thorough, albeit impressionistic descriptions of AAE prosody. Tarone's data showed greater variation in final tones/boundary tones in Vernacular Black English (VBE, i.e. AAE), than in Vernacular White English (VWE, i.e. MAE). Tarone found that there was a greater use of high pitch at phrase boundaries in VBE, and VBE used more level final contours for general questions while the VWE group used more rising contours. Hudson & Holbrook's (1981) study involving measurement of the reading fundamental frequency for African-American speakers provided further evidence of the frequent use of high pitch in AAE.

Jun & Foreman (1996) provide the first in-depth quantitative and qualitative analysis of AAE based on the model of intonation proposed by Pierrehumbert and her colleagues (Pierrehumbert 1980, Beckman & Pierrehumbert 1986). They found that AAE speakers use a wider pitch range and higher pitch, particularly at phrase boundaries, than MAE speakers. The study also revealed a categorical difference in boundary tones between AAE and MAE speakers for Yes-No questions and Declaratives. For genuine Yes-No questions in MAE, the stressed syllable of the pitch accent word has a low tone (L*), followed by a high phrasal tone (H-) and an unstepped high boundary tone, often with a sharp, but short incline at the tip (H%). Though AAE speakers did make use of the standard H-H% contour for Yes-No questions, they often use a high flat tone (H-L%). AAE speakers also used H* L-L% patterns in their Yes-No questions a pattern more commonly associated with Declaratives in both dialects. MAE declaratives and Wh-questions are characterized by a high tone (H*) marking the nuclear pitch accented syllable of the phrase, followed by a low phrasal tone (L-) and then a low boundary tone (L%). For Declarative and Wh-Questions, there is less degree of variability between the two dialects than what is found in Yes-No questions.

The documentation of these differences in intonation patterns between dialects called for studies to determine whether these measurable differences could also be perceived. A study by Lass, et al. (1979) addressed the issue of just how much of a signal was necessary before intonation differences could be perceived, assuming they were perceivable at all. They examined the effects of a phonetic signal of varying complexity (from isolated vowels to sentences) on the identification of speaker race and sex. Listeners were asked to identify the speaker’s race (Black, White) and sex. They found that more complex stimuli yielded more accurate judgments of speaker race, and that listeners were generally more accurate in their identification of White speakers than Black speakers.

Once it was ascertained that dialect-specific intonation differences were in fact perceptible to listeners, identifying the component of the signal that was most relevant for cue was another issue that needed to be addressed. Lass, et al. (1980) examined the effect of filtered speech on speaker race and sex identification. They found that accuracy in identification of race diminished with the application of a low-pass filter. Clearly these studies provided much needed information regarding methodology in terms of type of stimuli. However, generalizations about the ease of perception of AAE and MAE could not seriously be made without a study that included both Black and White listeners. By introducing this additional variable, one could determine to what degree experience with a particular dialect affected accuracy in identification of that dialect.

A study by Walton and Orlikoff (1986) served to fill the gap in terms of listener variability. Their listener pool included both Black and White listeners who were asked to identify the race of the speaker and to rate their level of confidence. What was most interesting was that those with exposure to AAE were more confident in their judgments than those without. However in this experiment, the stimuli consisted of a sustained vowel, not connected speech.

1.2. Purpose of the Study
The purpose of this study is to examine the use of intonational patterns in the identification of African-American English (AAE) and Mainstream American English (MAE). The study was developed to determine to what degree these intonational differences were used in distinguishing speaker dialect. Effects of exposure to AAE dialect on accuracy and confidence of identification was also examined. To do this, the listener groups were divided according to their exposure to AAE. It was predicted that the phrases with the strongest intonational distinctions found in Jun and Foreman (1996) would be most accurately identified.
2. EXPERIMENT

2.1. Materials
Perception data for this study is taken from the recordings made in Jun & Foreman (1996). In that study, 5 African-American pairs (four males and six females) and 2 White pairs (two males and two females) of speakers were recorded enacting scripted dialogues.

The White speakers were all members of the UCLA Linguistics Department and familiar with the sound booth procedures, and the African-American speakers were all visitors to the department. So although the script was geared towards AAE speakers, there was a certain level of discomfort for each group. The selected volunteer speaker pairs were friends, which contributed to the ease of interaction.

Fifty-four sentences were extracted from these recordings for this particular experiment based on distinctive intonational patterns. The distinctiveness could be in the area of dialect specificity, or in dialect mismatch, such that certain tokens within a dialect differed from each other. The recordings of the sentences were digitized using the Xwaves program, and waveforms and their corresponding pitch tracks were generated.

The sentences were then low-pass filtered at 900Hz cut-off frequency and 8kHz sampling frequency to mask obvious segmental and voice quality cues, while preserving prosodic cues. This was done by generating and applying a filter file within the Xwaves program. Due to variables such as clarity of speech and weakness of the signal, some tokens were more unintelligible than others.

2.2. Method
Listeners were asked to identify the speaker dialect for each sentence, and to rate their confidence for each decision using a 3-point scale (1 is high confidence, 2 is some certainty/uncertainty, 3 is no confidence). There was a 5-second pause between each token. Listeners were permitted to pause the tape manually or to listen to a token again if they needed to, but once they had passed an item, they could not go back. After listening to the filtered version, the answer sheet they used was removed, and then they listened to the unfiltered version, and rated again.

2.3. Subjects
A total of 39 listeners, 20 African-American (ten males and ten females) and 19 White listeners (nine males and ten females) were divided into 6 groups shown below, based on their race and their degree of exposure (from greatest to least) to AAE. The number in parentheses is the number of listeners in each group.

BC (9): Black code-switchers; This group uses AAE regularly, but is also able to use MAE.
B-W (8): Blacks who are not socially assimilated into White culture, but who are in an environment where MAE is spoken and who regularly speak MAE. They do not speak AAE, but they have an awareness of lexical, phonological and syntactic features of AAE.
W+B (3): Whites who are socially assimilated into Black culture on a regular basis during a significant portion of their lives ("My husband is Black" vs. "One of my neighbors is Black"). This group can speak AAE.
B+W (3): Blacks who are socially assimilated into White culture. Speakers of MAE, with no knowledge of AAE. W-B (8): Whites who are not socially assimilated into Black culture, yet have been in an environment where AAE was spoken.
W (8): Whites who have had basically no auditory exposure to AAE other than through mass media.

2.4. Hypotheses
Four hypotheses were set forth. First, dialect identification will be most accurate in those cases where the intonation pattern of the utterance matches the established obligatory tonal sequence for the sentence type in that dialect, thereby allowing for utilization of prosodic cues for identification. Second, in cases where the intonation pattern of a sentence does not match within dialect group (i.e., given two AAE speakers saying the same sentence with different intonational patterns), accuracy will be diminished for one or both of these sentences. Third, the degree of listener's exposure to African-American culture will be positively correlated with the accuracy rate of identifying the African-American speakers. Fourth, those listener groups who are least familiar with AAE will have a lower level of confidence in their identifications as evidenced by greater usage of 3 as a rating of their confidence.

3. RESULTS and DISCUSSION

3.1. Listener Variables
As there were fifty-four sentences in the data set, the highest number of correct identifications possible was fifty-four. When the data was examined with number of errors as the constant, the value for "mean error" was used. The mean error is the average number of errors for a given group of listeners. When the score was the constant, a value for "mean score" was used. Mean score represents the average score for a given group of listeners.

3.1.1. The Effect of Listener Sex and Race. An ANOVA with score plotted as a function of listener sex was performed to determine the interaction between the two variables. The results showed no significant interaction between score and listener sex (F(1, 37) = 3.065, p = .986). The mean score out of a possible 54 was 45.55 (standard error, 1.02) for the females. For males the average score was 45.53 (standard error, .88). Females rated their confidence at 3 for an average of three utterances out of the fifty-four. For males the average was two out of fifty-four utterances. So although the average scores were nearly identical for both males and females, females showed less confidence in their identifications.

As for listener race, the mean score for Black listeners was 46.55 (standard error, .91) or 86.2%, and the mean score for White listeners was 44.48 (standard error, .94) or 82.4%. Though the Black listeners performed slightly better, the difference in mean scores was not significant (F(1, 37) = 2.516, p = .121). There was however a notable difference between listener groups in their level of confidence in dialect identification. Black listeners rated their confidence at level 3 for an average of 1.8 utterances. White listeners rated their confidence at level 3 an average of 3.5 utterances. Low confidence rating from White listeners may be a function of their hesitancy to commit themselves to what they perceive as a racial stereotype.

3.1.2. The Effect of Exposure on Overall Performance. Figure 1 shows how each group performed on the 54 sentences. The general trend of the scores support a direct relationship between exposure to AAE and performance on the identification task. The only deviation from predicted performance lies with
groups W+B and B+W whose scores, based on their social exposure were expected to have been in reverse order.

3.1.3. The Effect of Exposure on Performance for AAE Sentences. An ANOVA was conducted with mean error as the independent variable, and exposure group as the dependent variable, to determine the effect of exposure to AAE on the identification of AAE tokens (Figure 2). The hypothesis predicting a positive correlation between exposure to African-American culture and accuracy in identifying AAE was supported by the data, though less strongly than expected. Though members of the African-American social circle made fewer errors overall, there was not as strong a correlation between exposure and accurate dialect identification as expected. Group W had the highest average number of errors (1.69, standard error, .31), and was followed closely by group W-B (1.66, standard error, .31). Group BC had a mean error of 1.34, (standard error, .19), B-W a mean error of 1.16 (standard error, .25). Groups B+W and W+B had the fewest errors with mean errors of .50 (standard error, .12) and .44 (.11 standard error), respectively.

Instead of those groups with the most exposure to AAE (BC & B-W) doing the best, it was those groups with the greatest amount of exposure to both cultures (B+W & W+B) that performed best. Identification accuracy for AAE actually decreased in accordance with listener affiliation with one particular culture. These results suggest that in order to effectively identify cues from AAE, one must be able to rule out MAE on the basis of one's familiarity with the features of that dialect. A greater number of participants in groups W+B and B+W would provide even stronger evidence in support of this.

3.1.4. Correlation between Exposure to AAE and Confidence. The incidences of confidence rating "3" was averaged to determine the distribution of usage. As predicted, group W used the level 3 confidence rating more often than any other group (average 4.38 incidents of use), followed by W-B (3.88) and B+W (3.00). Group BC, with an average of 1.78 incidents, used this rating more frequently than both B-W (1.38) and W+B (.33). It appears that there was is a strong overall relationship between exposure to AAE and confidence in identification of all sentences, such that those exposed to AAE on a regular basis were more confident in their decisions than those who are not thus exposed. Furthermore, within the group of listeners exposed to AAE, the hierarchy was shifted in favor of those who have more exposure to MAE.

3.2. Speaker Variables

3.2.1. The Effect of Prosodic Cues. In order to determine whether or not the presence of a prosodic cue made a difference in the accuracy of dialect identification overall, an ANOVA was performed with mean error as a function of cue for each dialect. The presence of cue increased the number of accurate responses for both AAE and MAE data. The mean error was slightly lower for the cued AAE tokens (1.02, standard error, .10), than it was for the non-cued AAE tokens (1.88, standard error, .38). However, the mean error for the cued MAE tokens (.63, standard error, .09) was substantially lower than that for non-cued tokens (2.61, standard error, .39).

That is, the absence of cue had a more detrimental effect on MAE identification than on AAE identification. More errors were made on non-cued MAE tokens than were made on non-cued AAE tokens.

Sentences spoken by an African-American were mislabeled more often than those spoken by a White American. There was a mean average of 1.13 errors overall for AAE sentences (standard error, .09), and an average of .90 errors overall for MAE sentences (standard error, .11). This was not surprising since African-American speakers do not always use the full range of AAE dialect features and the intonation pattern for each phrase type is not always dissimilar (i.e. Wh-questions).

Pitch range alone was also used as a cue for identification, but was not as useful as dialect-specific pitch patterns. Those tokens with both tonal sequence and pitch range cues generally had a high rate of accurate identifications.
3.2.2. Identification Accuracy for Dialect Mismatches. The prediction for cases where the intonation pattern of a sentence does not match within dialect group was supported by the data, though not without exception. Out of the six sets of sentence tokens that fell under this category, there were four that supported the hypothesis, and two that did not. Each token set consisted of a particular sentence being spoken in each dialect at least once. In some cases there were two speakers from a particular dialect group. In the latter cases, the intonation pattern for the speakers of the same dialect did not match each other. Only one of the two tokens fit a dialect-specific pattern in terms of pitch range and pitch contour (as established in [3]). The other token either matched for only one of the cues or else it resembled the prosodic pattern of the other dialect. The tables below show the sentence, its dialect, the number of errors made in identifying the sentence, and provides information about the type of cues available for a particular token set.

Table 1: Dialect mismatch between AAE tokens.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Dialect</th>
<th># of Errors</th>
<th>Cue Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeah he out here</td>
<td>AAE</td>
<td>13 errors</td>
<td>matches MAE</td>
</tr>
<tr>
<td>Yeah he out here</td>
<td>MAE</td>
<td>6 errors</td>
<td>range, contour</td>
</tr>
<tr>
<td>Yeah he out here</td>
<td>AAE</td>
<td>12 errors</td>
<td>range, contour</td>
</tr>
</tbody>
</table>

The same phenomenon occurs with the MAE data. In Table 2, none of the tokens matched another for tonal pattern. Crucially, the first MAE token did not match the second one, and AAE did not have a recognizable dialect-specific pattern. The result was a higher incidence of misidentification than that which is found in Hypothesis 1 situations.

Table 2: Dialect mismatch between MAE tokens.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Dialect</th>
<th># of Errors</th>
<th>Cue Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>I know you didn’t just say Ricky.</td>
<td>MAE</td>
<td>14 errors</td>
<td>range</td>
</tr>
<tr>
<td>I know you didn’t just say Willy</td>
<td>AAE</td>
<td>9 errors</td>
<td>Non-cued</td>
</tr>
<tr>
<td>I know you didn’t just say Ricky.</td>
<td>MAE</td>
<td>7 errors</td>
<td>range, contour</td>
</tr>
</tbody>
</table>

Table 3: Correct dialect identification of MAE without prosodic cues.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Dialect</th>
<th># of Errors</th>
<th>Cue Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baby got stupid body</td>
<td>MAE</td>
<td>2 errors</td>
<td>Non-cued</td>
</tr>
<tr>
<td>Baby got stupid body</td>
<td>AAE</td>
<td>2 errors</td>
<td>range, contour</td>
</tr>
</tbody>
</table>

One plausible explanation is that there were uncontrolled non-segmental cues such as voice quality that provided information regarding speaker dialect.

4. CONCLUSION

The overall high scores support the claim that prosodic cues are indeed a significant factor in identifying speaker dialect. The results show that performance is to some degree inversely proportional to the degree of exposure to both cultures, so that the more exposure one has to both cultures, the better one is at identifying the dialects. As one moves towards either end of the cultural spectrum, identification accuracy diminishes. This is a very logical assumption, and should be considered in future studies involving listener judgments of ethnic dialects. In addition, further research on dialect identification would contribute much to this area of study by examining the issue of voice quality, as it appears to be the final unexplored territory of dialect distinctiveness.

ACKNOWLEDGMENTS

My grateful thanks go out to the students at UCLA and Stanford's Ujamaa House, and to my friends and neighbors who participated in this study.

NOTES

1 AAE is a variety of English spoken with systematic phonological, syntactic and intonational patterns associated with members of the African-American culture. MAE is a cover term for that which is often called Standard English. It is not associated with a particular ethnic group, or a particular region (such as Southern).

REFERENCES