5aSC2. Applying pattern recognition to formant trajectories: a useful tool for understanding African American English (AAE) dialect variation

Meisam K. Arjmandi, Laura Dilley, Zach Ireland

Michigan State University, Department of Communicative Sciences and Disorders, East Lansing, MI

Abstract

Surprisingly few studies have carefully investigated the acoustic-phonetic characteristics of African American English (AAE) that distinguish this dialect from Standard American English (SAE), particularly for vowels and sonorant consonants. We investigated whether formant dynamics from short, sonorant portions of speech are sufficient to distinguish AAE and SAE dialects. Seven female speakers, four SAE and three AAE, from the Lansing, Michigan area were selected from a corpus of 30-45 minute sociolinguistic interviews. Target portions of speech consisting of a V or VC sequence (with C = /n/, /m/, /l/, /r/) were identified from contexts selected to control for coarticulation. F1 and F2 were extracted from randomly selected tokens at points 19%, 56%, and 81% of the duration through the demarcated speech portions. Pattern recognition techniques differentiated tokens of the two dialects based on formant trajectories as feature vectors. The results revealed that formant dynamics of some contexts are acoustically informative enough to differentiate the SAE and AAE dialects. These findings highlight the usefulness of incorporating pattern recognition techniques for understanding acoustic variation due to dialect.

Background

- Listeners can rapidly identify racial background only from the word *hello* (Purnell et al., 1999). However, the acoustic-phonetic characteristics that underlie these identifications remain unknown. Discrimination in housing options and in medical, judicial, and educational settings can result from racial dialect identification (Purnell et al., 1999; Baugh, 2000; Rickford & King, 2016).
- Many of the approximately 45 million African Americans speak African American English (AAE). Syntactic characteristics of AAE are well-described (Fasold & Wolfram, 1972; Baugh, 2000). However, little research has examined acoustic-phonetic characteristics of AAE, including vowel and voice properties (Kreiman & Sidtis, 2011; Thomas, 2007; Morris, 1997), with most work on F0 and consonants (e.g., Morris, 1997; Xue & Fucci, 2000).
- Formant trajectories are more informative than static formants (Morrison & Nearey, 2013). Formant dynamics from some specific contexts are potentially a major source of dialect identification.

- **Research questions:**

  1) Do formant dynamics provide reliable acoustic correlates to distinguish African American English (AAE) from standard American English (SAE)?

  2) Do some sounds and phonological contexts provide more reliable acoustic information for distinguishing AAE and SAE?

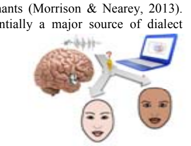

Figure 1. Acoustic-phonetic correlates provide the basis for auditory recognition of AAE and SAE dialect. Machine learning is a useful tool for studying what these correlates might be.

Methods

Analysis of Formant Trajectories:

- Talkers were three female AAE speakers and four female SAE speakers from Lansing, MI who completed sociolinguistic interviews. Tokens of vowels conditioned on specific phonological contexts were identified:
  - Closed syllables with a sonorant coda (/l/, /r/, /n/, or /m/) or non-sonorant coda, from a specific lexical items, to control for coarticulation.
  - For example, we extracted all instances of root morpheme *bark* to examine /ɑr/ and all instances of root morpheme cap to examine /æ/.
- We analyzed as many tokens as were available of the phonological context in question, up to a maximum of 10 randomly selected tokens of each phonological context for each speaker. Analysis of speech waveforms and spectrograms to measure F1 and F2 for each token from three time points (19%, 56%, 81%). The formant values were extracted and hand-corrected by trained analysts.

Pattern Recognition for Dialect Separation:

- Different classifiers were then trained on F1 and F2 trajectories; ***Feature set =*** $\{F1_{t1}\ F2_{t1}\ F1_{t2}\ F2_{t2}\ F1_{t3}\ F2_{t3}\}$. Classifier training was done on a subset of available tokens; these trained classifiers were then utilized to distinguish the dialect (AAE/SAE) from the unseen test set (Fukunaga, 2013).

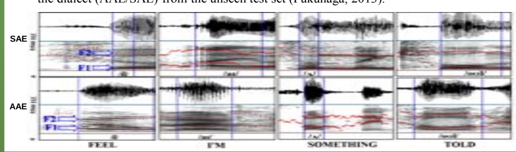

Figure 2. Formants are represented by red dots on the spectrogram. The series of figures in the top row are from a SAE speaker and the figures in the bottom row belong to an African American English talker.

Discussion and Conclusion

- The present study focused on acoustic properties that may allow listeners to identify racial background from pronunciation cues for AAE vs. SAE, which is an unsolved problem in phonetics (Purnell et al., 1999; Kreiman & Sidtis, 2011).
- These results show that pattern recognition and machine learning techniques can be usefully applied to determine which acoustic-phonetic properties may differentiate two dialects.
- The results showed that formant trajectories of F1 and F2 are highly distinctive and readily classifiable by machines as distinct, but only for a subset of phonological contexts. This suggests that certain speech sounds and phonological contexts are likely more informative to listeners about racial background than others.
- Individual variation in the degree of dialect realization is apparent for talkers of SAE and AAE.
- The present research aims to shed light on factors which are tied to bias experienced by talkers of a non-standard dialect, AAE, in everyday life, including educational, medical, and legal settings (Baugh, 2000; Rickford & King, 2016).

References

Baugh, J. (2000). *Beyond Ebonics: Linguistic pride and racial prejudice*. (Oxford University Press).

Fasold, R. & Wolfram, W. (1972). Some linguistic features of Negro dialect. *Language, Speech, and Hearing Services in Schools,* 3, 16-49.

Fukunaga, K. (2013). Introduction to statistical pattern recognition. Academic press.

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.

Morris, R. J. (1997). Speaking fundamental frequency characteristics of 8-through 10-year-old white-and African- American boys. *Journal of Communication Disorders*, 30(2), 101-116.

Morrison, G., & Nearey, T. (2013). Vowel inherent spectral change. (Springer Science & Business Media).

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.

Rickford, J. R. & King, S. (2016). Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language* 92, 948-988.

Thomas, E. R. (2007). Phonological and phonetic characteristics of African American vernacular English. *Language and Linguistics Compass*, *1*(5), 450-475.

Xue, S. A., & Fucci, D. (2000). Effects of race and sex on acoustic features of voice analysis. *Perceptual and motor skills*.

Acknowledgements

We would like to thank Dr. Suzanne Wagner's Sociolinguistics Lab at Michigan State for providing the corpus of interviews. We would also like to thank Alex Mason. We acknowledge Rachel Jansen and Chase Smitterberg and members of MSU Speech Perception-Production Lab for their assistance in data analysis.

Results

Pattern of Formant Trajectories Across Two Dialects of AAE and SAE

- Results indicated clusters in F1-F2 space for AAE and SAE talkers for multiple contexts.
- One SAE talker (KT) often patterned with the AAE talkers, consistent with findings that a talker's race can sometimes be misidentified perceptually (Tucker & Lambert, 1969). These findings support a hypothesis that racial dialect is learned, rather than due to biologically heritable properties of vocal tracts.

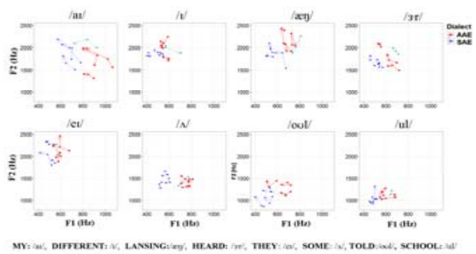

Figure 3. Formant trajectories of 8 different vowels for all speakers. Clusters among dialects are represented by the groupings of differently colored trajectories. Blue and red lines respectively represent formant trajectories of SAE and AAE in F1-F2. KT, an SAE speaker, is represented in teal because her pronunciations did not pattern with the other SAE talkers.

Pattern Recognition for Dialect Separation

- High accuracy across classifiers was found for multiple sonorant contexts.
- *F1* was more informative than *F2* for classifying certain sounds, while the reverse was true for others.
- Dialect separation is not feasible from some contexts (near-chance accuracy)

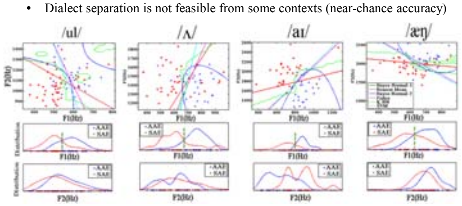

Figure 4. The distribution of F1 and F2 for four different phonological contexts. Top half: The performance of seven classifiers' decision boundaries are illustrated. Bottom half: Probability density functions of F1 and F2 are shown. The dialect separation ability from F1 and F2 varies among different contexts. Some contexts show greater degrees of separation than others.

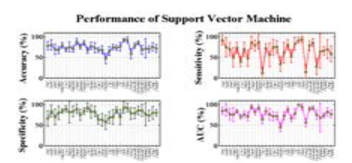

Figure 5. Results for AAE vs. SAE differentiation achieved by the SVM classifier.

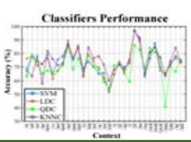

Figure 6. The classifiers' performance is shown in various contexts. Some contexts show greater accuracy, indicating greater differentiability for those contexts between dialects.