Perceptual sensitivity to dialectal and generational variations in vowels

Robert Allen Fox, Ewa Jacewicz

Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, USA.

fox.2@osu.edu, jacewicz.1@osu.edu

Abstract

Perception of dialect variation is well studied with respect to perceptual similarity of talkers based on dialectal markers. This study examines the perceptual distinctiveness of regional vowel variants in light of cross-generational changes in vowel productions. Listeners from two regional dialects of English identified the dialect of the speaker in monosyllabic words (produced by older adults, young adults and children). Differential listener sensitivity to speaker dialect was found, which was highly affected by speaker generation. This suggests that the ability to determine dialect membership is an interaction between the perceptual spaces of listeners and the acoustic variations in vowels.

Index Terms: dialect variation, vowel perception, sound change

1. Introduction

Regional variation in the vowel system of American English is well documented in sociophonetic literature and summarized in the seminal *Atlas of North American English* [1]. Most work in this area has used relatively basic acoustic measurements of vowel properties (such as midpoint formant frequencies and vowel duration) to explicate the scope of differences in vowel productions across regional dialects in the United States [2]. Recently, new insights into the nature of phonetic variation have been brought to light in a number of studies which utilize a large multi-speaker corpus and provide a more detailed description of time-varying spectral changes in regional vowel variants. In particular, speakers from the North (Wisconsin) displayed a dispersion pattern in the acoustic vowel space and dialect-specific formant movement that was distinct from that in the Midlands (central Ohio) and in the South (North Carolina) [3]. The nature and amount of formant dynamics were found to vary not only across dialects but also within each regional variety as a function of speaker generation and gender [4, 5]. Furthermore, the Northerners were found to have the shortest vowel durations from all three dialects [6], perhaps related to their faster speech tempo as compared to the slower articulation rate of Southern speakers [7].

A natural question arises as to the perceptual distinctiveness of these regional vowel systems. That is, given the significant cross-dialectal variation in the way vowels are pronounced, to what extent is the acoustic information in vowels sufficient to allow listeners to recognize and/or identify regional accents? Recent research in perceptual classification of regional varieties of American English focused on several dialect markers such as r-fullness, fricative voicing, vowel brightness, vowel backness and degree of diphthongization [8, 9]. These studies found that, in general, listeners can make accurate judgments about both the identity and the similarity of these regional varieties. However, in these tasks, experimentally constructed read sentences were used to elicit the perceptual response and listeners may have used a variety of cues (and not only the specific markers) to categorize talkers in terms of their dialectal background.

Another indication that listeners use their knowledge of and experience with dialectal features comes from lexical processing studies which also measured reaction times. For example, the results of [10] indicate that experience with a dialect has a major effect on the ability of listeners to process variants which contained r-ful and r-less forms. There was a consistent processing cost for speakers hearing variants from a non-native dialect, which was evident both in recognition and lexical activation. In a different study which examined processing costs involved in regional accent normalization in French [11], it was found that an unfamiliar accent increases reaction times, causing a delay in word identification in continuous speech. This and related research increases our understanding of how language comprehension system adapts to regional accent variation and underscores the role of experience in the recognition and representation of dialects.

The present study examines listener sensitivity to regional variation in vowel productions within and across two varieties of American English spoken in western North Carolina (NC) and southeastern Wisconsin (WI). Acoustic characteristics of these two regional vowel systems indicate that each of them is currently undergoing a set of changes which affect both positions of vowels in the acoustic space and the amount of formant movement [4]. These changes are manifested in speech of each younger generation from the same geographic area. The question we are addressing now is whether long-time residents of each geographic area are sensitive to cross-generational differences in vowels produced by older adults, younger adults and children and can recognize these variants as belonging to their own dialect. That is, when exposed to a high level of cross-generational variation in the input containing exemplars from their own and a different dialect, do NC listeners "know" which variants are from their own NC dialect and which are from the WI dialect? Conversely, are WI listeners sensitive to this type of variation and can they recognize which variants belong to their own regional variety?

1.1. Cross-generational variation in vowels

Dynamic vowel dispersion patterns in the formant 1 by formant 2 acoustic vowel space for three generations of female speakers from NC are shown in Figure 1 and from WI in Figure 2. The three age groups are operationally labeled grandparents (GP), parents (P) and children (C). In order to illustrate generational differences in vowel patterns in each of these two dialects, we have selected and graphed a subset of vowels /i, æ, ɑ, o, u, aɪ/ that were produced in the isolated tokens *heed, had, hod, hoed, who'd, hide* (more details on stimulus materials are provided in the next section). For each vowel, lines connect five measurement points from 20% to 80% and vowel symbols are placed next to the 80% point to indicate the direction of formant movement. The mean formant frequency values shown in the figures are based on the subset of tokens from our large speech corpus used in the perception test and are provided here as an illustration of cross-generational variation in the two regional vowel systems. To reduce effects of the vocal tract length differences

between adults and children, the formant values were converted to *z*-scores using Lobanov's normalization [12].

Among the NC speakers, we find that GP speakers have a centralized variant of /i/ which also has a comparatively greater formant movement than in the P and C speakers, producing the phonetic percept of a diphthongized vowel. The vowel /i/ is produced somewhat more fronted in P and C compared with GP. Another cross-generational change can be seen in /æ/. This vowel is produced progressively lower in the acoustic space with each younger generation and the nature of its formant movement is different in C. The diphthong /aɪ/ is produced basically as a monophthong in both sets of adult speakers but as a diphthong by children. The vowel /ɑ/ is raising in the acoustic space across generations and has comparatively greater formant movement in C. The vowel /o/ is produced farther back in C and /u/ is produced as a very fronted variant in all three generations (more retracted in C).

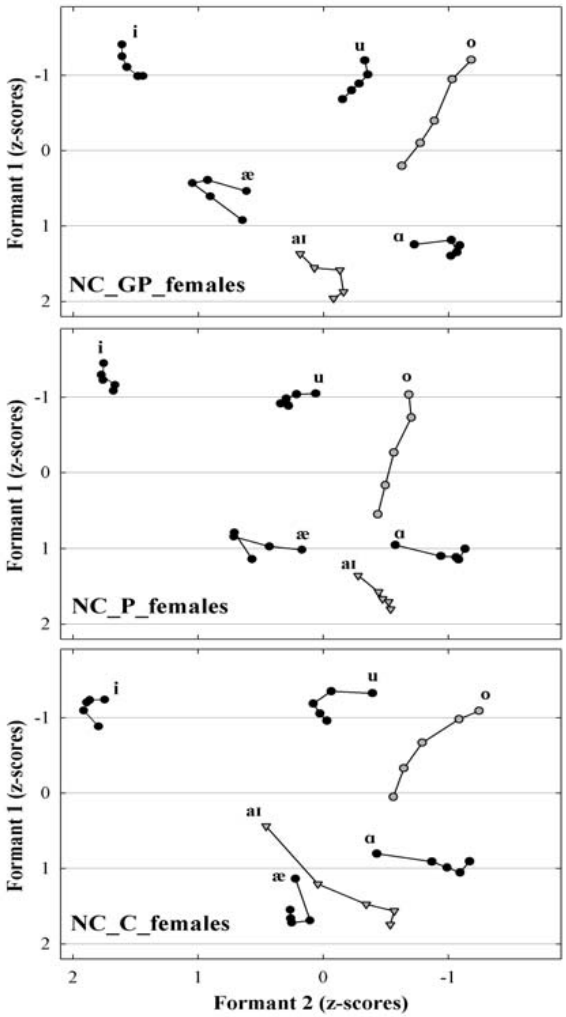

Figure 1: *Cross-generational variation for selected North Carolina vowels.*

The WI variant of /i/ does not change its fronted position across generations. The vowel /æ/ has a different extent and direction of formant movement compared to NC variants and is relatively stable across WI generations (although some backing and lowering of /æ/ can be found in C). WI /aɪ/ is a true diphthong which is in sharp contrast with the NC variants (especially in GP and P). WI /ɑ/ is produced as a very low variant in GP which occupies the position of NC monophthongal /aɪ/. As can be seen, WI /ɑ/ is produced in a higher and more retracted position in P and C. Finally, the WI variants of /u, o/ are monophthongal and produced far back and in a close proximity to one another in GP. Although not changing its back position, the P and C variants of /o/ have a greater amount of formant movement although still substantially reduced compared to NC variants of /o/.

We hypothesize that these types of generational and dialectal variation will elevate the level of stimulus uncertainty for both NC and WI listeners. A perception test was conducted to verify their ability to identify these two dialects and to examine how well they can cope with cross-generational changes in vowel quality (in terms of both position and spectral dynamics).

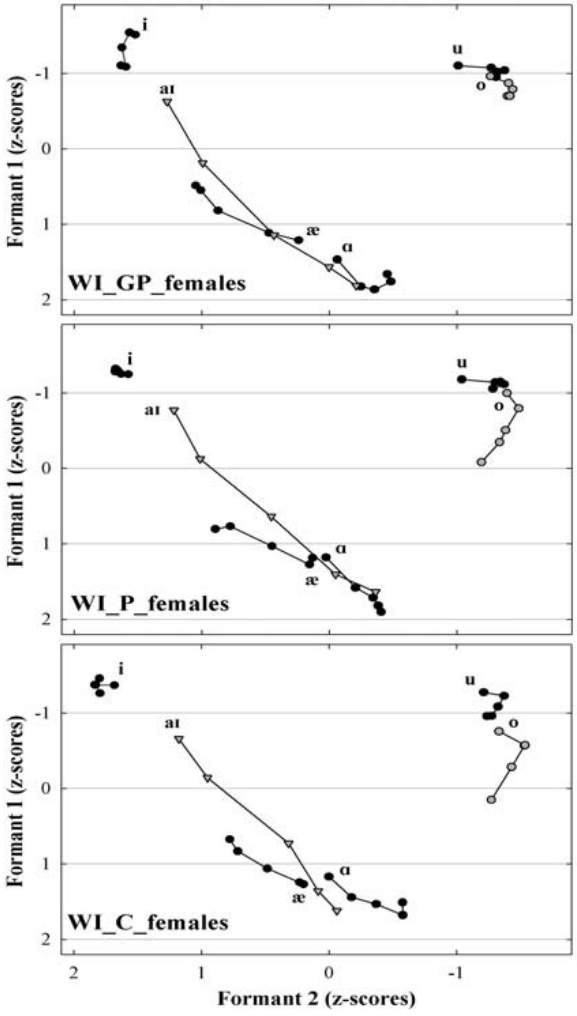

Figure 2: *Cross-generational variation for selected Wisconsin vowels.*

2. Methods

2.1. Listeners

Listeners were 30 adults who were born, raised and spent most of their lives in the same regions as the speakers, i.e., in either western NC (15 listeners) or southeastern WI (15 listeners). Their ages ranged from 43-58 years. Each listener spoke a

local regional variety of American English typical of each geographic area, i.e., a version of Southern English in NC and a version of Midwestern English in WI, as verified by research staff. They were mostly professionals recruited through local community announcements and none of them participated in any speech experiments before. All listeners reported normal hearing and none of them was phonetically trained.

2.2. Stimulus material

Natural speech was used in the experiment. The stimuli consisted of hVd tokens *heed, hid, hayed, head, had, hod, who'd, heard, hide, hoed, hood, hawed* containing 12 vowels: /i, ɪ, e, ɛ, æ, ɑ, u, ɝ, aɪ, o, ʊ, ɔ/. The tokens were recorded as isolated words (one at a time) and were selected from a large corpus of recordings completed for a production study of regional variation in American English vowels [4]. For the present experiment, the productions of 120 speakers whose ages represented three generations of long-time residents in each dialect region were chosen at random (60 in NC and 60 in WI). In each dialect, the tokens were produced by 20 children (C = 8-12 years old), 20 young adults (P = 35-50 years old) and 20 old adults (GP = 66-91 years old). There were 10 male and 10 female speakers in each age group.

Six different token types from each speaker (half of the set of 12) were selected for a total of 720 unique exemplars for use in the perception experiment (6 tokens x 20 speakers x 3 generations x 2 dialects). Acoustic measurements (obtained using LPC analysis with a 25-ms Hanning window) included frequency of the first three formants (F1, F2 and F3) sampled at the 20-35-50-65-80% points in the vowel (to estimate dynamic formant movement) and vowel duration (more details can be found in [3]). These measurements were used to verify cross-generational differences in vowel productions.

2.3. Procedure

The experiment was conducted in a quiet room at two locations, at Western Carolina University in Cullowhee, NC, and University of Wisconsin-Madison, WI. The same experimental set-up and experimental protocol was followed at both sites. Signals were delivered over headphones and each listener was tested individually. During a one-hour session, each listener was presented with 720 signals in three randomized blocks of 240 tokens each. Prior to the experiment, a 20-item practice was administered to each listener to familiarize him or her with the task.

Upon hearing a single signal, listeners responded by clicking with the mouse on one of two boxes on the computer screen which displayed two possible responses "Wisconsin" and "North Carolina." The stimulus presentation and response collection was under the control of a custom Matlab program. The listeners were told that they would hear one word at a time and had to decide whether the word was produced by a WI or a NC speaker. No more than one repetition was allowed and the listeners were instructed to guess if they were not sure which response to choose.

2.4. Data analysis

In order to examine the ability of listeners to identify the dialect of a speaker, we converted the identification responses to d-prime (d') values [13]. Correct WI responses to WI tokens were "hits" and incorrect WI responses to NC tokens were "false alarms". These d' values (calculated not only for the total sets of responses, but for individual vowels, speaker subsets and listener groups) provide a measure of sensitivity to dialect variation while eliminating response bias.

3. Results

In the analyses below, we report d' values, a measure of sensitivity to the dialect difference. This measure is preferable to percent correct dialect identification as it removes response bias. In the first set of analyses, we examined overall sensitivity of either NC or WI listeners to all 720 exemplars presented for dialect identification. NC listeners were most sensitive to GP productions, followed by P and C, respectively (d' values were 0.96, 0.83 and 0.45, respectively). WI listeners were comparatively less sensitive to GP speakers but rated P and C speakers similarly as NC listeners (d' values were 0.83 0.80 and 0.52, respectively). As a reference, higher d' values denote greater sensitivity; typical values are up to 2.0 and 69% correct for both NC and WI trials corresponds to a d' of 1.0.

The second set of analyses examined listener sensitivity to speaker dialect as a function of individual vowels and the age (i.e., generation) of the speaker. In order to conserve space in this paper, we present only the results for the six vowels /i, æ, ɑ, o, u, aɪ/ in *heed, had, hod, hoed, who'd, hide* (the same vowels that were shown for female speakers in Figures 1 and 2). The sensitivity data for NC listeners are displayed in Figure 3. The highest sensitivity to tokens produced by all three generations was for the vowel /u/ in *who'd* and the lowest was for /aɪ/ in *hide*. For the latter, NC listeners were unable to correctly identify the dialect of the speaker when the tokens were produced by either children or P adults. Low scores were also obtained for /ɑ/ in *hod*, although there was still some dialect sensitivity to the exemplars produced by GP and P adults. The scores for the remaining three vowels /i, ɑ, o/ show relatively high sensitivity to GP exemplars, somewhat lower to P and the lowest to C productions (except for /o/). Altogether, these results indicate that the cross-generational differences among vowels resulting from sound change over time do significantly influence listeners' decisions about the dialect of the speaker.

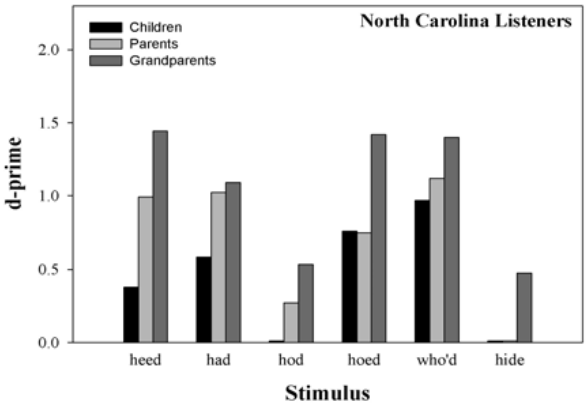

Figure 3: *Sensitivity scores for NC listeners.*

The sensitivity data for WI listeners are shown in Figure 4. As can be seen, the overall pattern of responses for individual vowels for these listeners differs noticeably from that of the NC listeners. The most drastic difference was found for /aɪ/. Unlike NC listeners, WI listeners were very sensitive to cross-dialectal differences, particularly when listening to vowels produced by P adults. Dialect sensitivity was lower for children's productions, suggesting that the diphthongal /aɪ/ in NC children (at variance from the monophthongal /aɪ/ in NC adults) was frequently undistinguishable from that in WI children, which shows again listeners' sensitivity to cross-

generational vowel change. WI listeners' identification of dialect membership was very good when listening to GP exemplars of /æ/.

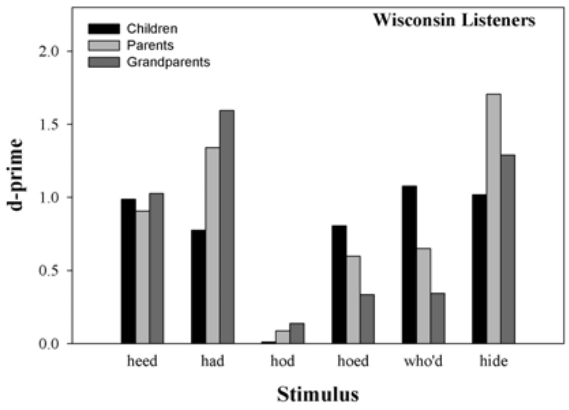

Figure 4: *Sensitivity scores for WI listeners.*

An interesting finding can be seen in the d' values for /o, u/, demonstrating the highest sensitivity of WI listeners to children's variants and lowest to GP adults. This pattern is clearly different compared to NC listeners and is not easily interpretable on the basis of acoustic characteristics of vowels in terms of their positional differences in the acoustic vowel space or the amount of formant movement. Unlike for NC listeners, the d' values for the vowel /i/ are similar across the three generations, showing that WI listeners can detect dialectal differences relatively well apart from speaker generation. Finally, WI listeners were basically insensitive to dialectal and cross-generational differences for /ɑ/ in *hod*, which parallels the finding for NC listeners although the latter could still detect some dialectal differences in the productions of adults.

4. Discussion and conclusions

Typically, experimental work on dialect recognition has employed single sentences or stretches of connected speech; it is therefore notable that in the present study, untrained listeners demonstrated dialect sensitivity while hearing a single monosyllabic word. Clearly, the acoustic characteristics of the vowel (formant values, spectral change and duration) were the prime contributors to the dialect identification decisions. Given that regional vowel systems of American English undergo sound change over time, the listeners were asked to respond to a mixture of "older" and "newer" pronunciation patterns of a given vowel category produced by male and female speakers of different generations. As is evident from Figures 3 and 4, sensitivity to such cross-generational variations differs as a function of listener dialect. This leads us to the conclusion that the ability to determine dialect membership is an interaction between the dialect-specific structure of the perceptual vowel spaces of the listeners and the acoustic differences in the vowels in the two dialects. Evidence from work on the perception of vowels in a second language would certainly be applicable here [14].

The results also indicate that different vowel categories provide different amount of information for the listeners to make the dialect distinction. We may not exclude here the role of experience with phonetic differences between the dialects. For example, the present WI listeners showed lower sensitivity to adults' variants of *who'd* despite a drastic phonetic difference between the highly fronted NC /u/ and a far back WI /u/. This would indicate their knowledge of variations in the amount of /u/-fronting, possibly through interactions with individuals from different dialect regions. Sensitivity of NC listeners' showed just the opposite, which would suggest their unfamiliarity with the WI variant. This possible explanation needs yet to be verified in a separate experiment, however.

At present, we are just at the beginning of understanding how listeners use the acoustic cues in vowels to make their choices about dialect identification. Listeners' sensitivity to cross-generational variation in vowel production informs us that acoustic characteristics are an important component in making these decisions in addition to other strategies they may be using. Future work will explore the complex relationship between the acoustic structure of vowels and listener sensitivity to dialectal distinctions.

5. Acknowledgements

This work was supported by NIDCD/NIH research grant R01DC006871. We thank Joseph Salmons for his contributions to this research and Janaye Houghton and Dilara Tepeli for help with data collection.

6. References

[1] Labov, W., Ash, S., and Boberg, C., *Atlas of North American English: Phonetics, phonology and sound change*, Berlin, Germany: Mouton de Gruyter, 2006.

[2] Clopper, C., Pisoni, D., and de Jong, K., "Acoustic characteristics of the vowel systems of six regional varieties of American English," Journal of the Acoustical Society of America, 118: 1661-1676, 2005.

[3] Fox, R., A., and Jacewicz, E. "Cross-dialectal variation in formant dynamics of American English vowels," Journal of the Acoustical Society of America, 126: 2603-2618, 2009.

[4] Jacewicz, E., Fox, R. A., and Salmons, J., "Cross-generational vowel change in American English," Language Variation and Change, 23: 45-86, 2011.

[5] Jacewicz, E., Fox, R. A., and Salmons, J., "Regional dialect variation in the vowel systems of normally developing children," Journal of Speech, Language and Hearing Research, 54: 448-470, 2011.

[6] Jacewicz, E., Fox, R. A., and Salmons, J., "Vowel duration in three American English dialects," American Speech, 82: 367-385.

[7] Jacewicz, E., Fox, R. A., and Wei, L., "Between-speaker and within-speaker variation in speech tempo of American English," Journal of the Acoustical Society of America, 128: 839-850, 2010.

[8] Clopper, C., and Pisoni, D., "Some acoustic cues for the perceptual categorization of American English regional dialects," Journal of Phonetics, 32: 111-140, 2004.

[9] Clopper, C., and Pisoni, D., "Perceptual similarity of regional dialects of American English," Journal of the Acoustical Society of America, 119: 566-574, 2006

[10] Sumner, M., and Samuel, A., "The effect of experience on the perception and representation of dialect variants," Journal of Memory and Language, 60: 487-501, 2009.

[11] Floccia, C., Goslin, J., Girard, F., and Konopczynski, G., "Does a regional accent perturb speech processing?" Journal of Experimental Psychology: Human Perception and Performance, 32: 1276-1293, 2006.

[12] Lobanov, B. "Classification of Russian vowels spoken by different speakers," Journal of the Acoustical Society of America, 49: 606-608, 1971.

[13] Macmillan, N., and Creelman, C., *Detection theory: A user's guide (second edition)*, Mahwah, NJ: Lawrence Erlbaum, 2005.

[14] Fox, R. A., Flege, J., and Munro, M., "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," Journal of the Acoustical Society of America," 97: 2540-2551, 1995.