The /ay/s Have It
THE PERCEPTION OF /AY/ AS A NORTH-SOUTH STEREOTYPE IN UNITED STATES ENGLISH

by

Bartłomiej Plichta and Dennis R. Preston
Michigan State University

Abstract
There is no doubt that /ay/ monophthongization is one of the principal caricatures of US speech. Few studies, however, have made use of phonetic detail in investigating regional or group identification and caricature, and none has addressed the possibility that a perceptual continuum parallels the continuous nature of some phonetic variables. This study attempts to determine if the degree of /ay/ monophthongization is relevant to perception along a north-south line of nine voices stretching from Saginaw, Michigan to Dothan, Alabama, the same used in an earlier study of dialect recognition by the second author. Some characteristics of both the speaker (sex) and hearer (sex, region) are also explored.

1. Introduction
Recent advances in speech science allow us to ask increasingly sophisticated questions about production and perception (e.g., Graff, Labov and Harris 1983, Labov and Ash 1997, Purnell, Idsardi, and Baugh 1999, Niedzielski 1999, Thomas 2001). In this paper we ask how sensitive hearers are to degrees of the monophthongization of /ay/ in United States (US) English. /ay/ monophthongization is one of the principal caricatures of southern US speech (Thomas 2001: 34); it is frequently imitated and cited in folk accounts of the speech of the south, itself the most distinctive linguistic area of the US for nonlinguists (e.g., Niedzielski and Preston 2003: 99-101). We want to confirm hearer sensitivity to this feature and explore whether a greater degree of monophthongization can be associated with deeper southern-ness. In asking this question we approach a more general one: Can speakers use degree of a continuous feature in variety identification? Since our question is sociolinguistic as well as dialectological, we also want to know if speaker and listener demographic information plays any role in this task.

2. Previous Studies
We need to know at first about actual speech divisions along the north-south continuum which we will use in this research and whether there is any evidence
that hearers distinguish this continuum in general. Based on previous perception research (e.g., Preston 1989a), we will use the north-south dimension shown in Figure 1.

![Map of US English dialect areas](image)

Figure 1. Nine sites indicating a north-south dimension in US English (Preston 1989a: 349)

The evidence from Linguistic Atlas studies, which investigate lexicon, grammar (principally morphology), and pronunciation, shows nine dialect areas along the continuum of sites from Figure 1. Figure 2 shows these nine divisions — A-C-D-Southern Indiana-Western Kentucky-1-K-P-R. The sites from Figure 1 are
distributed in these Atlas studies as follows: Saginaw and Coldwater are in A, South Bend is in C, Muncie is in D, New Albany is in Southern Indiana, Bowling Green is in Western Kentucky, Nashville is in I, Florence is in K, and Dothan is in R.

Figure 2. Linguistic Atlas regions of the Eastern United States (Lance 1994: 352).

Unfortunately, the data on which many of the divisions in the northern half of this map are based come from as yet unpublished work on LANCS (the Linguistic Atlas of the North-Central States), a part of the Linguistic Atlas of the United States and Canada, and the major studies of this area derived from those data focus more on vocabulary than pronunciation (e.g., Dakin 1971, Marekwardt 1957), making it difficult to say where /ay/ monophthongization begins. Although data from roughly the southern half of Figure 2 are readily
available (in LAGS, the *Linguistic Atlas of the Gulf States*, Pederson 1986-92), there is no indication that this feature subdivides the large southern region in any way except for those smaller areas in which /ay/ monophthongization occurs before voiceless consonants rather than only before voiced consonants and in final position.

The same north-south continuum according to The Phonological Atlas of North America (Labov, Ash, and Boberg, to appear) shows only four subdivisions: Inland North-North-Midland-South (Figure 3).

![Figure 3. The Phonological Atlas of North American English (Labov, Ash and Boberg, to appear, Figure 11.13)](image)

If /ay/ monophthongization is a clue to the north-south dimension, however, this work suggests only a two-way distinction, since Labov et al. use that feature as the identifying characteristic of “The South,” with a northernmost boundary running right on top of New Albany, IN (see Figure 1). As in LAGS, the identifying characteristic of the “Inland South” in Figure 3 is the extension of /ay/ monophthongization to an environment before voiceless consonants, a conditioning factor we will ignore in this study.
It has not been the case, however, that nonlinguists' mental maps of regional speech areas necessarily correspond to these of actual production. Preston (e.g., 1989b) has proposed a variety of techniques to determine folk perceptions of dialect areas, and since this study will focus on nonlinguists' responses, they are also reviewed here.

In one technique, respondents are asked to draw and label the speech areas of the United States on a blank map. Figure 4 is an example of such a hand-drawn map.

![Hand-drawn map of US speech regions by a South Carolina respondent](image)

Figure 4. A hand-drawn map of US speech regions by a South Carolina respondent (Preston 1996: 311)

These individual maps have been converted by computational techniques into a general account of a group of respondents' mental dialect maps. Figure 5 shows such a generalization for 147 southeastern Michigan respondents. If the north-south dimension of Figure 1 is overlaid on this map, there are four divisions — North-Midwest-Inner South-South, a very similar distribution to Figure 3, which tempts one to believe that folk perceptions are at least in part phonetically based.
Figure 5. A southeastern Michigan mental map of US speech regions (Preston 1989a: 344)

In this representation, the first four sites of the continuum of Figure 1 (Saginaw to Muncie) are contained in the North; New Albany is in the Midwest, Bowling Green and Nashville are in the Inner South, and Florence and Dothan are part of the South.

Figure 6 shows another mental map, derived by similar techniques, for 123 southern Indiana respondents. Again, there are four clear divisions (North-
Midwest-Inner South-South), but the boundaries are in different places. Only Saginaw and Coldwater are in the North; South Bend and Muncie are in the Midwest, and New Albany is at the border of the Midwest and the Inner South. Bowling Green and Nashville are in the Inner South, and, as in the Michigan representation, Florence and Dothan belong to the South.

In both these studies of folk perception, there is the suggestion that nonlinguists recognize four distinctive areas along this north-south line.

Figure 4, however, and many other hand-drawn maps by respondents, suggests that the task of drawing speech area boundaries by nonlinguists does not take place in a world in which purely linguistic matters are considered. If that were so, one would not find such labels as “Northern Scratch & Claw” and “South Courteous & Gentlemanly & Also Spoken by Ignorants” (Figure 4). Another clue to such a prescriptivist view can be found even in the frequency with which areas are identified. Of the southeastern Michigan respondents (Figure 5), 0.94 draw a boundary around a southern speech area; the next most frequently outlined region is the local northern area at 0.61, and the next is an area with its center in New York City (0.54). Similarly, the southern Indiana respondents outline the south most frequently (0.86) and the New York City area second (0.51). One need not study a large number of hand-drawn maps to determine the source of this salience. The south and the New York City area are most frequently singled out as areas where incorrect English is spoken, and the upper Midwest or north (where Michigan is located) is the area most frequently identified as the home of correct English.

The hand-drawn maps, therefore, did not reveal directly what respondents believe about the regional distribution of English in the US; they did, however, reveal how these respondents made use of their principal folk theory of language – language correctness (e.g., Niedzielski and Preston 2003: 96) – in carrying out what, for a dialectologist, would have been a purely geographical task. This folk theory of correctness in US English, and, doubtless, elsewhere, is an important part of the investigation of language ideology:

By foregrounding ideology I emphasize the need to investigate ideas about language and speakers independently of empirical distributions, and the need to recognize that “attitudes” include participants’ basic understandings of what the sociolinguistic system consists of, not just emotional dispositions. Moreover, the categories and behaviors toward which one has these attitudes cannot be assumed to have been established independently of anyone’s perceptions of them (Irvine 2001:24)
Figure 7. Southeastern Michigan geographical ratings of correct English (Preston 1989a: 333)

Figure 8. Southern Indiana geographical ratings of correct English (Preston 1989a: 332)
To assess the prescriptive foundation of regional speech distinctions more directly, the Michigan and Indiana respondents were asked to rate the fifty states, Washington, D.C., and New York City on a scale of 1 to 10 (1 = least correct, 10 = most correct) for language correctness. Figures 7 and 8 show the results of this task. Figure 7 shows that southeastern Michiganders have a robust view of the distinctiveness of correct English by region. They descend five full steps on the correctness scale from their home (Michigan, the most darkly shaded area at the top of the map, the most correct, in the 8.00 – 8.99 range of mean scores) to Alabama (least correct, in the 3.00 – 3.99 range). The southern Indiana respondents (Figure 8), however, find only two degrees of correctness along the scale under investigation here – a generally northern 6.00 – 6.99 and a generally southern 4.00 – 4.99.

We must be prepared, therefore, if these correctness caricatures influence perceptions of a phonetic element, such as /ay/ monophthongization, to encounter quite different responses from respondents from different geographical areas.¹

Next, we consider the few studies which have actually asked respondents to match voice samples to regions. Generally, dialect recognition has not been particularly good. Both Williams, Garrett, and Coupland (1999) and Clopper and Pisoni (2004) found only thirty percent accuracy when respondents were asked to match regional voices to home areas, although Van Bezooijen and Gooskens (1999) found higher rates for both British and Dutch dialects. Only two studies, however, have focused on identification which includes or focuses on the US north-south dimension being studied here.

Clopper and Pisoni (2004) played speech samples (with appropriate regional phonetic features, e.g., New England r-lessness) for Indiana University students. The voices were from New England or the West (areas which do not concern us here) or from the North (north of the Michigan-Indiana boundary), the North Midland, (the upper one-third of Indiana) the South Midland (the lower two-thirds of Indiana) or the South (everything south of the Ohio River, i.e., the Indiana-Kentucky boundary). The sites from Figure 1 are, in these terms, distributed as follows: Saginaw and Coldwater are North, South Bend is North Midland, Muncie and New Albany are South Midland, and everything else (Bowling Green through Dothan) is South.

¹A second dimension, that of language “pleasantness” has been studied (e.g., Preston 1989a), but those results are not studied here since the distinctions for the southeastern Michigan respondents is very similar to the results for their correctness rankings, and the results for southern Indiana show only a two-way distinction (i.e., Indiana is very pleasant, other areas are not as pleasant).
As noted above, correct identification of voices in the Clopper and Pisoni study was low, about thirty percent overall, but the assignment of voices to sites was not random. Cluster analyses showed that significant groupings, with minor variation among respondents from different regions, linked the South and South Midland as one area, the North, North Midland, and West as another, and New England stood apart. That grouping suggests that a major boundary lies between South Bend and Muncie (see Figure 1), a boundary supported by Figure 3, although the other major boundary of Figure 3 (between New Albany and Bowling Green) is not.

Clopper and Pisoni did not find /ay/ monophthongization to be a significant part of their South or South Midland speaker speech samples, so it does not turn up in their careful regression analysis of the relationship between their speech samples’ phonetic features and the respondents’ regional assignments. Their focus on the word *like*, which, as they note, has a final voiceless segment, the more expanded environment for monophthongization, may have been the source of this failure.

Finally, we turn to an earlier study (Preston 1989a) which played actual speech samples from the sites shown in Figure 1 and asked respondents, again

<table>
<thead>
<tr>
<th>Michigan Judges</th>
<th>Indiana Judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Mean</td>
</tr>
<tr>
<td>1 Saginaw</td>
<td>3</td>
</tr>
<tr>
<td>2 Coldwater</td>
<td>5</td>
</tr>
<tr>
<td>3 South Bend</td>
<td>1</td>
</tr>
<tr>
<td>4 Muncie</td>
<td>2</td>
</tr>
<tr>
<td>5 New Albany</td>
<td>6</td>
</tr>
<tr>
<td>6 Bowling Green</td>
<td>4</td>
</tr>
<tr>
<td>7 Nashville</td>
<td>8</td>
</tr>
<tr>
<td>8 Florence</td>
<td>7</td>
</tr>
<tr>
<td>9 Dothan</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 1. Ranks of regional voices, north to south, with mean scores derived from site numbers assigned each voice by the respondents (based on the sites shown in Figure 1); both Indiana and Michigan assignments show overall significance (ANOVA p < .00001).
from southeastern Michigan and southern Indiana, to identify which site each voice was from. The nine sample voices were all of young to middle-age male speakers, all with some college education. The samples were drawn from sociolinguistic interviews, and all lexical and grammatical clues to region were avoided. Table 1 shows the results of this task.

Both the Michigan and Indiana respondents assigned site numbers (from Figure 1, in which the sites were numbered “1” [Saginaw] through “9” [Dothan]) to the voices they heard, allowing the mean scores to be ranked from highest (northernmost) to lowest (southernmost). As Table 1 shows, these scores display a strong southern identification but considerable confusion in the placement of northern and middle voices.

Although ANOVA runs for both Michigan and Indiana voice placements were significant, the post-hoc (Tukey) analyses of paired differences for each voice was complex, and the relations are better shown in cluster analyses (Figures 9 and 10). For both groups of judges, there are two major areas of distinctiveness; there is a clear affiliation of the three southernmost sites (Nashville, Florence, and Dothan) as opposed to everything else, as shown by the immediate combination of these three areas on the leftmost side of the cluster analyses in both Figures 9 and 10. In both, however, there is considerable variation in the placement of voices in the northern and middle regions. Both Indiana and Michigan respondents, however, add Saginaw, the northernmost site, to the large northern-middle group last (i.e., rightmost in the cluster analyses), suggesting that they may have perceived it as more distinctly northern than any other area. Both the New Albany and Bowling Green voices belong to the large northern-middle group, in spite of the fact that both show quite strong evidence of /ay/ monophthongization.

This actual dialect placement test is somewhat like Figure 3 and suggests again that nonlinguists are sensitive to phonological matters in dialect identification, a conclusion also supported by Clapper and Pisoni (2004). This study goes on to look at a specific feature, the degree of /ay/ monophthongization, which may contribute to the folk ability to distinguish positions along the north-south continuum shown in Figure 1. It is important to remember, however, that we use these degrees of monophthongization only as clues to the use of such a continuous feature in stereotyping by our respondents. As the survey of earlier studies above shows, we have no evidence of either different frequencies of monophthongization or degrees of it in the areas where it occurs.

2 These data have been presented several times in earlier publications. Unfortunately, in Preston 1996a and Niedzielski and Preston 2000, the figures are incorrect for both Michigan and Indiana, and there are minor errors in Preston 1993 and earlier presentations of the same research.
3 Background
Such work in general involves a large number of questions, only a few of which we will approach in the present study:

a. Will small acoustic modifications in the monophthongization of /ay/ be perceived?
b. If perceived, will the degrees of monophthongized /ay/ be perceived gradiently or will there be some (or several) categorical break-points?
c. Will speaker characteristics (age, sex, etc...) affect perception?
d. Will hearer characteristics (age, sex, region, ethnicity, etc...) affect perception?
Our approach will be to present seven resynthesized versions of /ay/, ranging from fully diphthongal to fully monophthongal, to hearers who will then select one of the nine sites in Figure 1 as being the one which that version of the word “guide” (which avoids the distributional difficulty with /ay/ before voiceless segments mentioned above) came from. We go on now to show the speech science background to this work and the preparation of both the stimuli and the website setting in which this experiment was conducted.

We first address the matter of categorical perception. Numerous speech perception studies have found rather poor discrimination of within category acoustic cues, though vowels in even across-category tests seem to be perceived slightly less categorically than consonants (Pisoni 1973). Based on the premises of the categorical perception of speech sounds and poor within-category discrimination, we might expect a very simple regional identification in our task; we may find that there is only a dichotomous division into Northern-sounding [ә] and Southern-sounding [әː], or we may find that the regional identification responses are near chance and that the respondents are not able to distinguish among the seven steps at all.

4. Methodology
The LPC analysis/re-synthesis method was employed in the preparation of the stimuli — a seven-step continuum of monophthongization of the phoneme /ay/ in the word guide, spoken by a male and female voice. The analysis was carried out pitch-synchronously (auto-correlation over a Blackman window), and the filter coefficients were computed independently for each pitch period. F1 and F2 values were computed for each frame at each of the seven degrees of monophthongization. Since there is evidence that fronting (F2) also accompanies monophthongization (e.g., Kurath and McDavid 1961), a minor adjustment was also made to that parameter. The maximum changes were 150Hz in F1 and 550Hz in F2; F3 remained unchanged. LPC residual representation of the glottal source was obtained through inverse filtering, which was then re-synthesized with the altered LPC coefficients. Figure 11 shows seven modified F1 and F2 dimensions (and the unmodified F3) for the male voice.

To carry out a study that would involve subjects from various parts of the United States, it became evident that it would be best to design a web-based experiment. The task was made available at http://bartus.org/ai (login: experiment, password: poland) and consisted of the following blocks: introduction and loudness calibration, consent form, demographic questions, practice run (3 trials), main block (randomized 42 trials – each male and female token at each of the seven steps repeated 3 times).

Finally, a map was given on the website so that respondents could indicate what part of the US they were from. The subdivisions were an attempt to
capture the important boundaries found in both production and perception studies of variation (see Table 3).

![Formant frequency vs. Time graph]

Figure 11. Formant tracks (F3 top, F2 middle, and F1 bottom) for the seven steps of the word *guide*. The most monophthongized version is represented by the bottom line in F2 and the top line in F1.

5. Results
The main results given in this report are based on scores which were assigned each hearing of a stimulus word (i.e., one of the seven steps of the word *guide*) on the basis of the values assigned the areas shown in Figure 1. If, for example, a respondent heard step 4 and associated it with New Albany, IN, then a score of ‘5’ was assigned and so on. The mean scores for each step in the monophthongization continuum are listed in Table 2, with map positions and their values shown to the right. An ANOVA post hoc test (Tukey) shows that each score is different from each adjacent score (and all others), and this robust differentiation shows that the minimal differences provided in the seven steps of /ay/ monophthongization can be used by folk respondents to discriminate along the north-south dimension indicated by the sites.
Table 2. Mean scores based on regional values assigned each step of the increasingly monophthongized versions of /a/

<table>
<thead>
<tr>
<th>Step</th>
<th>Mean</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.85</td>
<td>1. Saginaw</td>
</tr>
<tr>
<td>2</td>
<td>3.17</td>
<td>2. Coldwater</td>
</tr>
<tr>
<td>3</td>
<td>3.87</td>
<td>3. South Bend</td>
</tr>
<tr>
<td>4</td>
<td>4.89</td>
<td>4. Muncie</td>
</tr>
<tr>
<td>5</td>
<td>5.99</td>
<td>5. New Albany</td>
</tr>
<tr>
<td>6</td>
<td>6.58</td>
<td>6. Bowling Green</td>
</tr>
<tr>
<td>7</td>
<td>7.02</td>
<td>7. Nashville</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8. Florence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9. Dothan</td>
</tr>
</tbody>
</table>

The next feature in addition to the steps themselves which proved significant was associated with the stimulus voice. As Figure 12 shows, there was a difference in the ratings of male and female voices at every step of the monophthongization continuum. The female voice was always rated farther north, and the difference is significant at every step (based on independent t-tests).

Figure 12. Differences in regional (north-south) placement of male and female voices
Table 3. Respondent frequency by region and sex

<table>
<thead>
<tr>
<th>Region &amp; Sex</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Pacific Northwest</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>B Northern CA</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>C Southern CA</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>D NV-UT</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E AR-NM</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F Plains &amp; Mountains</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G CO-NB-KS</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>H TX</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>I MN</td>
<td>11</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>J WI &amp; MI</td>
<td>3</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>K North Midland</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>L South Midland</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>M Gulf South</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>N New Orleans</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>O Upstate NY</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P Pittsburgh</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q Central PA</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>R NYC</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>S New England</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>T Philadelphia</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>U DC</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>V Eastern VA &amp; NC</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>W GA &amp; SC</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>X South FL</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Y Hawaii</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Z Alaska</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Totals:</strong></td>
<td><strong>35</strong></td>
<td><strong>61</strong></td>
<td><strong>96</strong></td>
</tr>
</tbody>
</table>
Although other significant effects will emerge from respondent identities, this finding for sex of stimulus voice is an especially interesting one for sociolinguists, for it shows, we believe, that, in spite of acoustic evidence, respondents are less willing to associate a female voice with a region which is more stigmatized, i.e. the US south (e.g., Preston 1996). This finding suggests that the long-standing association of women’s speech with more standard languages practices (e.g., Trudgill 1972) is evident in this perception task, although we are well aware of the complexities of this generalization (e.g., Eckert and McConnell-Ginet 1992).

We turn our attention now to various aspects of respondent characteristics as they influence placement of the seven steps. Table 3 shows the number of respondents by geographical area and sex who have responded so far. It is clear that the number of respondents in each of these regional groups is insufficient for individual statistical treatment, and we have conflated them to four, as follows:

- West = A, B, C, D, G (N=16)
- South = H, L, M, V, W, X (N=22)
- North = I, J, K (N=38)
- East = F, Q, R, S, T, U (N=20)

The letters refer to the sites listed in Table 3. Table 4 shows respondent age.

<table>
<thead>
<tr>
<th>Age</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-25</td>
<td>27</td>
</tr>
<tr>
<td>26-40</td>
<td>30</td>
</tr>
<tr>
<td>41-60</td>
<td>35</td>
</tr>
<tr>
<td>61-77</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>96</td>
</tr>
</tbody>
</table>

Ethnicity is ignored in the following since only a very few respondents (7) did not classify themselves as European Americans (1 African American, 2 Asian Pacific Islanders, 1 Hispanic, and 3 who did not disclose ethnicity). ANOVA tests for significance were run on the male and female stimulus voice samples independently since the placement of the steps was shown to be influenced overall by sex of speaker (Figure 11). The independent variables tested were 1) sex of hearer, 2) regional background of hearer (based on the four groupings detailed above), and 3) age group of hearer (Table 4).
Age group of hearer was rarely selected as significant and is not included in any of the following discussion.\(^3\)

Sex of hearer was also not selected as significant except in one case out of the possible 14 — step #3 for female voice (which males rated as significantly more “southern” — 3.50, compared to the female rating of 3.10, t-test, \(p = .03\)). Hearer sex is also, therefore, disregarded in the following.

The four regions into which the respondents were sorted, however, were significant at every step for male voices except for #7 and were significant for a few steps (#2, #4, #6) for female voices. These regional differences in evaluation are shown in Figures 13 (male voice) and 14 (female voice).

![Graph showing regional assignment of male voices based on respondent background.](image)

**Figure 13.** Assignment of male voices based on regional background of respondent

\(^3\) A more robust analysis was also done with a regression analysis on actual age of respondent but was still not significant.
As Figure 13 shows, there is a clear preference among southern and western respondents to rank male voices as more southern when there is no (step #1) or relatively little (steps #2 and #3) monophthongization. In contrast, the Midwestern respondents are the most reluctant to rank a male voice as southern, regardless of the degree of monophthongization. These findings add interesting corollaries to the overall perception of female voices as less southern. Figure 14, by contrast, shows a tendency for easterners and southerners to agree on female voices as more southern (step #s 2, 4, and 6), although at very different levels of monophthongization. In fact, without the recognition of female voices with advanced degrees of monophthongization by easterners as southern, the overall tendency for female voices to be rated as more northern would be even greater. Further investigation of the regional distribution of hearers may reveal that some of the patterns uncovered here are more affected by region than these data might show, perhaps even revealing patterns of interaction between region of respondent and sex of speaker.

A not unexpected finding in this experiment was that there was apparently no awareness of any of this discriminating ability among the respondents themselves. Although, as noted above, speakers of US English often comment on or even imitate /ay/ monophthongization as a southern (or African American) speech stereotype, the respondents did not believe that they were capable of
discerning among the degrees of monophthongization presented here. Most felt it was simply an impossible task, one which would yield no interesting results. Comments such as the following two were very common.

I don't do surveys ordinarily. I tried this one only because of my great respect for you, as evinced from your contributions to the oral history netserve. You have been a great help to me. Unfortunately, after doing the sample, I realized that I am clueless. My ignorance could only compromise your findings. Had I a printed version of the word before it was sounded, I might have had more of a snowball's chance.

I've just done the web linguistic survey and I kept going even though I was quite clueless. How should anyone other than travelling salesmen have any familiarity with these regional dialects? Who's a "good respondent" as opposed to one like me who's mostly guessed in a tripartite way: north, middle, south? I had one stop in Indianapolis once during which I was so struck by a sixteen year waitress's accent that I've always remembered the difficulty of understanding her, but I sure haven't remembered her voice.

I hope your survey produces something of value, but I'm curious how you might get there if other respondents are as clueless as I am!

We comment further on respondent cluelessness below.

6. Discussion

On the speech science side, it is perhaps not surprising that these modified steps are perceivable in terms of the acoustic distance from one step to the next, but it is perhaps phonologically surprising that some sort of within-category categorical perception is not the outcome of the task. The finely-tuned overall discrimination is robust.

From a perceptual dialectological point of view it would appear that continuity rather than discreteness in boundaries is possible. Respondents appear to classify things as "more southern" or "more northern" rather than "north" or "south" (or "north", "midland," "south"), even when there is no evidence in production studies that such an acoustic continuum exists. This respondent treatment of such a variable may allow even more sensitive attitude and discrimination studies of dialect and sociolect variation than previously thought worthwhile.
In the area of sociophonetics, stereotypes (here of sex) clearly influence perception, but it is not surprising that social information may overwhelm acoustic detail. Niedzielski (1999), shows, for example, that Michigan respondents are unwilling to hear non-canonical vowels when they believe the speakers are fellow Michiganders, an unwillingness which she attributes to the strong Michigan belief that their speech is standard or correct (e.g., Figure 7). Strand (1999) shows that typical male versus female identity can modify a hearer’s perception of sibilant consonants.

In this study, female voice almost certainly awakens in the hearer the sociolinguistic commonplace that women are typically speakers of more standard varieties than men (e.g., Labov 1990) and that in self-report women exaggerate the standardness of their speech while men exaggerate their own nonstandardness (Trudgill 1972). Male voice, on the other hand, leads southern and western respondents to find male voices without the clue of monophthongization to be more southern. This, we believe, exploits the solidarity side of gender identity with regard to region. Female voice is associated with more northernness to distance it form the perceived nonstandardness of southern speech; male voice is heard as more southern, in effect as more masculine, a quality associated with more vernacular speech (e.g., Trudgill 1972). In short, we find that the caricatures of north-south correctness combines with those of male-female standardness to deny the acoustic clue of an equally southern speaker (on the basis of degree of /ay/ monophthongization) to a female voice sample and to allow the placement of a voice without this acoustic clue to be judged as more southern so long as the voice is male.

Finally, it is clear that respondents have discriminatory abilities which far outstrip their conscious awareness. Even in presentations of this seven-step continuum to audiences of linguists, many trained in phonetics, a common reaction is that the differences between adjacent steps are not detectable. These contrary findings would seem to confirm Preston’s (1996b) idea that availability, accuracy, detail, and control are separate clines or continua in the levels of awareness as regards folk knowledge of language. The feature /ay/ is clearly available to all US English speakers (e.g., they discuss it), and they have a fairly accurate representation of it (i.e., they know that monophthongization is the element involved). It is a detailed fact about southern English (not a global one), and many speakers have some control over it in imitation. When subjects are presented with degrees of monophthongized /ay/, however, nearly every one of these generalizations about the overt or folk view disappears. Gradience is not available to them overtly, although they clearly respond to it; they do not accurately represent their sensitivity to gradience overtly, however, and no caricaturistic imitations we are aware of ever involve more-or-less monophthongized varieties of /ay/.
7. Conclusions
As a result of this work we believe even more strongly that finely-grained acoustic differences are not only the basis for investigations of actual sound systems and changes in progress but may also be manipulated in experimental settings to confirm perceptual salience in such tasks as social category identification.

This work further confirms that social stereotypes (e.g., sex) may play an important role in the study of the perception of varieties, which, in turn, may be directly linked to variety acquisition trends.

Finally, it is clear that perceptual discrimination in such overt tasks may operate well below the conscious level.

References


**About the authors**

Dennis R. Preston  
Department of English  
e-mail: preston@msu.edu

Bartłomiej Plichta  
Department of Linguistics and Germanic, Slavic, Asian and African Languages  
e-mail: plichtab@msu.edu

Michigan State University  
East Lansing, MI USA 48824