Published in final edited form as:

J Phon. 2004 January 1; 32(1): 111–140.

Some acoustic cues for the perceptual categorization of American English regional dialects

Cynthia G. Clopper* and David B. Pisoni

Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405, USA

Abstract

The perception of phonological differences between regional dialects of American English by naïve listeners has received little attention in the speech perception literature and is still a poorly understood problem. Two experiments were carried out using the TIMIT corpus of spoken sentences produced by talkers from a number of distinct dialect regions in the United States. In Experiment 1, acoustic analysis techniques identified several phonetic features that can be used to distinguish different dialects. In Experiment 2, recordings of the sentences were played back to naïve listeners who were asked to categorize talkers into one of six geographical dialect regions. Results showed that listeners are able to reliably categorize talkers using three broad dialect clusters (New England, South, North/West), but that they have more difficulty categorizing talkers into six smaller regions. Multiple regression analyses on the acoustic measures, the actual dialect affiliation of the talkers, and the categorization responses revealed that the listeners in this study made use of several reliable acoustic–phonetic properties of the dialects in categorizing the talkers. Taken together, the results of these two experiments confirm that naïve listeners have knowledge of phonological differences between dialects and can use this knowledge to categorize talkers by dialect.

1. Introduction

Studies of phonological differences in regional dialects of American English have focused primarily on the collection of phonological atlases, phonological descriptions of specific dialects, or the social aspects of attitudes towards certain dialects, such as perceived “correctness” and social stereotypes related to speakers of a given dialect (e.g., Giles, 1970; Labov, 1972; Labov, Ash, & Boberg, in press; Preston, 1993). Linguistic atlas projects have been in progress in the United States since the early 1900s with the goal of documenting regional phonological and lexical variation (e.g., Kurath, 1939). Labov conducted some of the first smaller-scale variation research in the 1960s in his well-known study on the use of [ɹ] in New York City department stores. This new research approach proved instrumental in shaping the current field of variationist research (Labov, 1972). At around the same time, Giles and his colleagues were also conducting attitude judgment research in Britain on native speaker attitudes towards varieties of English spoken in the British Isles. They asked participants to listen to recordings of speech and rate the talker on dimensions such as aesthetics and status (e.g., Giles, 1970).

While dialect geographers have typically focused more on lexical variation than on phonological variation, early linguistic atlas projects collected information about both lexical and phonological variation (e.g., Kurath, 1939). More recently, Labov et al. (in press)

*Corresponding author: cclopper@indiana.edu (C.G. Clopper).

have been working on a complete phonological description of North American English, using data collected from telephone surveys of over 600 talkers across the United States and Canada. The recordings from these talkers were impressionistically transcribed and acoustic measurements of F_1 and F_2 were obtained for each of the vowels they selected to study. Based on the differences in vowel production, their *Atlas of North American English* identifies various levels of dialect boundaries that range from a basic North–South–West split to the division of New England into Eastern New England, Western New England, and New York City.

In another study, Thomas (2001) explored the vowel systems of close to 200 speakers of North American English using F_1 and F_2 acoustic measurements taken from recordings of read passages and spontaneous conversational speech materials. The bulk of the speakers in his study came from Ohio, North Carolina, or Texas, which allowed him to carry out a thorough analysis of the differences between these three regions, as well as provide a sense of the degree of variation within a given locale, such as a state or even a city. In his presentation and discussion of the vowel spaces of individual talkers, Thomas made two important contributions to the acoustic literature on variation. First, he provided a quantitative means of comparing individuals within and across dialect areas. Second, he used a single methodology in the analysis of every talker in his corpus, thus allowing for the direct comparison of materials that had previously been collected for disparate projects but had never been presented together in a standard format.

The collection of a large number of samples of spoken language has enabled researchers to study specific dialects. Similarly, the study of specific dialects has led to the collection of large spoken language corpora. In studying and describing individual varieties of American English, the focus has typically been on the vowel system of that variety (Docherty & Foulkes, 1999). The current shifts in the vowel systems of two regions in particular have received a great deal of attention in the past few decades: the Northern Cities vowel shift and the Southern vowel shift. The Northern Cities vowel shift is characterized by a “clockwise” rotation of the low vowels in the $F_1 \times F_2$ vowel space as shown on the left in Fig. 1. It has been found in such urban areas as Buffalo, Cleveland, Detroit, and Chicago (Labov, Yaeger, & Steiner, 1972). The Southern vowel shift, on the other hand, is characterized by the centralization of the high tense vowels and the lengthening of the high front lax vowels as shown on the right in Fig. 1. This shift is found more prominently in rural areas of the South, as opposed to the more urban populations that exhibit the Northern Cities vowel shift (Labov & Ash, 1997).

A third phenomenon involving vowels in American English that has received attention in the literature is the low-back merger in which /ɔ/ and /ɑ/ have merged to make homophones of such pairs as ‘caught’ and ‘cot’ or ‘Dawn’ and ‘Don.’ The quality of the merged vowel varies from talker to talker, and ranges from [ɒ] to [a]. This low-back merger is found in the Midland areas of the United States and much of the West (Wolfram & Schilling-Estes, 1998; Labov et al., in press). While vowels have been the primary focus of phonological dialect descriptions, consonantal phenomena like the postvocalic r-lessness found in New England, New York City, and some parts of the South, and the ‘greasy’ ~ ‘greazy’ alternation found in the South have also been noted dialect characteristics in discussions of phonological differences (Labov, 1972; Wolfram & Schilling-Estes, 1998).

In studying the perception of these phonological differences by naïve listeners, the methods that are typically used have been based on the representations of dialect variation that listeners have stored in memory, and not on direct behavioral responses to any speech stimuli. For example, Preston (1986, 1989) conducted a series of studies in which he asked undergraduates from various parts of the United States to complete a number of tasks,

including drawing and labeling dialect regions on a map of the United States and ranking all 50 states on the “correctness” or the “pleasantness” of the English spoken there. Studies were conducted in Hawaii, southern Indiana, eastern Michigan, New York City, and western New York. The results indicated that naïve undergraduates cannot accurately replicate the dialect boundaries drawn by such variationist researchers as Labov.

A comparison of the composite maps of each participant group indicated that concepts of dialect variation are related, in part, to where a listener lives. In general, listeners defined more dialect regions in areas that were in closer geographic proximity to themselves than in areas that were farther away (Preston, 1986). Similarly, results of the ranking task for informants in southern Indiana indicated that “pleasantness” seemed to correspond to geographic proximity to Indiana, whereas “correctness” seemed to correspond more to highly familiar stereotypes of where “standard” English is spoken, with California and the North and Northeast regions receiving the highest rankings (Preston, 1989). The results of these perceptual dialectology studies may have been affected by the participants’ poor knowledge of United States geography. Preston admitted that using a map without state boundaries resulted in great confusion and that “for the time being, folk dialectology is confounded with folk geography” (1993, p. 335). Similar problems in perceptual dialectology research have been noted in Great Britain (Inoue, 1999; Wales, 1999).

In addition to the concern about participants’ geographic knowledge, one major criticism of this research strategy is that the participants are rarely, if ever, asked to listen to actual speech samples when completing these perceptual tasks. Instead, they are asked to make judgments based on their mental representations of dialect variation stored in long-term memory. It is therefore unclear whether or not the participants could reliably identify any given talker as being from a place that they rate as having “pleasant” or “correct” English.

In addition to Preston’s (1986, 1989, 1993) perceptual dialectology work, a substantial body of literature exists that has examined attitude judgments based on listening to samples of spoken language. Lambert, Hodgson, Gardner, and Fillenbaum (1960) introduced a methodology known as the “matched-guise technique” in which the same talker produces utterances in multiple languages or dialects in order to control inter-talker variability while obtaining the desired inter-language or inter-dialect variability. Listeners are then typically asked to rate the talker on a series of scales related to intelligence, likeability, competence, trustworthiness, etc. Giles and his colleagues (Giles, 1970; Giles & Bourhis, 1973; Bourhis, Giles, & Lambert, 1975; Ryan & Carranza, 1975) have used this technique to explore linguistic attitudes of naïve listeners to a range of regional, ethnic, and social language varieties and have consistently found that talkers exhibiting features of less prestigious dialects receive lower ratings than talkers exhibiting features of more prestigious dialects.

More recently, Purnell, Idsardi, and Baugh (1999) assessed the ability of naïve listeners to identify the dialect of a talker, using a variation of the matched-guise technique. The male talker in the study by Purnell et al. left answering machine messages for landlords in various neighborhoods in the San Francisco area inquiring about apartments for rent using white, African-American, and Chicano guises. The results suggested that the landlords were able to identify the talker’s dialect (or guise) based on these short samples of speech because the number of returned calls for African-American and Chicano guises increased as the minority population of the neighborhood increased.

One major criticism of matched-guise research is that participants are rarely asked to identify where they think the talkers are actually from. Interpretations of the results therefore are often based on the assumption that the listeners first correctly identified where the talker was from and then completed the attitude judgment portion of the task. The validity of this

assumption, however, is questionable, particularly given other research in the domain of dialect perception. In his famous study of the linguistic attitudes of teachers in Chicago, Williams (1976) found that both white and African-American children who were identified as white received higher ratings on a number of language-related scales such as fluency, standard pronunciation, and sentence complexity than children who were identified as African-American. He concluded that the teachers may have made their attitude judgments based on their perception of the child's race, instead of the actual linguistic characteristics displayed by the child.

More recently, Niedzielski (1999) examined the ability of naïve listeners in Detroit to match synthetic vowel tokens to a target vowel spoken by a single female talker. One group of listeners was told that the talker was from Detroit. A second group of listeners was told that the talker was from Canada. Niedzielski found that those listeners who thought the talker was from Canada selected the actual matching vowels from the set of six alternatives. The listeners who were told that the talker was from Detroit, however, selected canonical vowels as the best match to the talker's utterances. That is, the perception of vowel quality was affected by the listeners' beliefs about the talker. As in Williams (1976), where the teachers' judgments were related to their perception of the race of the talkers, the listeners in Niedzielski's study were influenced by their stereotypes of Canadian vs. Detroit English. Taken together, these results suggest that an important area of research is the study of the identification of regional, ethnic, and social dialects by naïve listeners.

A small number of studies have been conducted that explicitly examine the ability of naïve listeners to identify where different talkers are from based on actual speech samples. In one of the earliest studies of its kind, Bush (1967) asked naïve listeners to identify the national origin of talkers from the United States, Great Britain, and India. She found that listeners could identify where the talkers were from with over 90% accuracy in a three-alternative forced-choice categorization task using read speech including nonsense words, real words, and sentences.

More recently, Preston (1993) asked naïve adult listeners in Michigan and Indiana to listen to short speech samples taken from interviews with middle-aged males and to assign each of the different talkers to one of nine cities, running north to south between Saginaw, Michigan and Dothan, Alabama. Results of his study revealed that the listeners were only able to make a broad distinction between North and South. Preston noted that this perceptual boundary did not correspond to the boundaries drawn by these same listeners in the map-drawing task discussed above, suggesting that listeners' perceptions of dialect variation when listening to actual speech samples differ significantly from their stored mental representations of dialect variation. It is also interesting to note that the boundary perceived by the Indiana residents was different from the boundary perceived by the Michigan residents. Preston (2002) suggested that the difference in categorization between the Michigan and Indiana listeners might be due to a different attentional focus. Specifically, he proposed that the Michigan listeners were making their identifications based on attitude judgments of relative "correctness," whereas the Indiana listeners were making their identifications based on judgments of relative "pleasantness." In any case, it was clear that the overall results of this identification task were similar across both groups, with slight inter-group variation due to where the listeners themselves were from.

Van Bezooijen and her colleagues (Van Bezooijen & Gooskens, 1999; Van Bezooijen & Ytsma, 1999) have examined perceptual dialect categorization of regional varieties of Dutch in the Netherlands and Belgium and English in the United Kingdom. Using speech samples taken from interviews with three male talkers from each of four regional varieties of Dutch, Van Bezooijen and Gooskens asked naïve Dutch listeners to identify the country, region,

and province that each talker was from in a multi-level forced-choice perceptual categorization task. They found that Dutch listeners could correctly identify 60% of Dutch talkers by region of origin and 40% by province. Van Bezooijen and Ytsma observed similar results using read passages spoken by four female talkers representing each of six Dutch varieties. Their listeners identified 60% of the talkers by region and 35% by province. In the United Kingdom, Van Bezooijen and Gooskens replicated their Dutch study with three male talkers from each of five English dialects and found that listeners identified 88% of the English talkers by region and 52% by area, again using speech samples taken from interviews.

Williams, Garrett, and Coupland (1999) recorded two adolescent males from each of six regions of Wales and two adolescent male speakers of Received Pronunciation (RP) recounting personal narratives. Short samples of these utterances were played back to different groups of adolescents from each of the six regions who were asked to categorize each talker using an eight-alternative forced-choice task. The eight alternatives were the six regions of Wales, RP, and “don’t know.” Compared to the results reported by Van Bezooijen and Gooskens (1999) and Van Bezooijen and Ytsma (1999), overall performance measured in terms of accuracy was quite low. The average proportion correct categorization was only 30%. In considering the performance of the listener groups, Williams et al. found that listeners only correctly identified talkers from their own region about 45% of the time.

The recent studies reviewed above on dialect identification and categorization based on actual speech samples have all found that listeners can make judgments about where talkers are from with some degree of accuracy. Given the varied nature of the studies themselves and how performance was scored, it is difficult to make direct comparisons among them. However, it is clear that Purnell et al. (1999) were tapping into the ability of landlords to identify a talker’s racial dialect based on an answering machine message, that the listeners in Preston’s (1993) study could discriminate northern from southern American English talkers based on short narratives, that the listeners in the Van Bezooijen and Gooskens (1999), Van Bezooijen and Ytsma (1999), and Williams et al. (1999) studies were able to categorize talkers by regional dialect with somewhat variable, although consistently above-chance performance. Taken together, these findings suggest that naïve listeners are aware of phonological differences between dialects and can make reliable judgments based on this information in the speech signal.

The present set of experiments extended this line of dialect categorization research in two ways. First, through the use of read speech materials that were identical across all of the talkers, we measured the acoustic–phonetic properties associated with the talkers from each of the dialects of American English included in our study. Second, using playback techniques, we carried out a perceptual categorization study of regional varieties of American English with American listeners that was similar to the research done by Williams et al. (1999). In addition to replicating their categorization results, we also conducted several detailed analyses of our perceptual data to investigate the nature of the errors made by our listeners and to measure the acoustic–phonetic properties that the listeners were relying on in making their categorization judgments.

The present experiments were designed to determine how naïve American listeners categorize talkers by regional dialect of American English and to identify the acoustic cues that are used by the listeners in making categorization judgments about where a talker is from. Wolfram and Schilling-Estes claim that “phonological patterns can be diagnostic of regional and social differences, and a person who has a good ear for dialects can often pinpoint a talker’s general regional and social affiliation with considerable accuracy based solely on phonology” (1998, p. 67). However, there is little, if any, experimental evidence

available in the published literature to show that listeners have this detailed knowledge of variation in phonological patterns at all, or that they can use this knowledge reliably as a diagnostic for regional identification. Thus, the primary goal of the present research was to investigate dialect variation in both speech production and speech perception. Experiment 1 was carried out to measure which acoustic–phonetic cues were available in the speech signal to identify where a talker was from. Experiment 2 was designed to investigate how listeners categorize talkers by dialect and describe which acoustic–phonetic cues are used by the listeners in making these perceptual categorization judgments.

2. Experiment 1: acoustic analysis of dialect variation

2.1. Talkers

Sixty-six talkers were selected from the TIMIT Acoustic–Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986; Zue, Seneff, & Glass, 1990). The TIMIT corpus consists of audio recordings of 630 talkers reading ten sentences each. The corpus includes 438 males and 192 females. The talkers were each assigned one of eight regional labels to indicate their dialect: New England, North, North Midland, South Midland, South, West, New York City, or Army Brat. While this corpus was initially designed for use in speech recognition research, it has also been used in a number of phonetic analyses that investigated the roles of gender, dialect, and age in language variation (e.g., Byrd, 1992; Keating, Blankenship, Byrd, Flemming, & Todaka, 1992; Byrd, 1994; Keating, Byrd, Flemming, & Todaka, 1994). Until the present study, the corpus had not been used in perceptual studies on dialect variation with human listeners.

The 66 talkers selected for the current phonetic study were white males between the ages of 20 and 29 at the time of recording. Eleven talkers were chosen from each of six dialect regions: New England, North, North Midland, South Midland, South, and West. Those talkers who did not meet the age, gender, and race requirements were eliminated for each of the six dialects. Given the limited nature of the demographic information provided with the TIMIT corpus, social factors such as socio-economic status and urban vs. rural distinctions could not be controlled or evaluated.¹

The eleven talkers from each region used in this study were selected by the first author and a second phonetically trained judge based on repeated listening to all of the materials available in the corpus for each talker. The talkers were selected such that they shared as many features of their dialect as possible. Specifically, all of the New England talkers that were selected for this study were “r-less” (i.e., non-rhotic). The Northern talkers were selected based on their degree of /æ/ raising and /oo/ fronting. The selected South Midland and Southern talkers produced monophthongal /aɪ/. Some Southern speakers also produced fronted /u/ or a merger of /e/ and /ɪ/ before nasal consonants. The Western speakers who were selected all produced fronted /u/ and some also produced the merger of /e/ and /ɪ/ before nasals or a merger of /ɑ/ and /ɔ/. Finally, the North Midland speakers selected for inclusion produced none of the characteristic features of any of the other five dialects.

2.2. Stimulus materials

Of the ten sentences spoken by each talker in the TIMIT corpus, two of the sentences were read by all of the talkers. These two “calibration sentences” were written to include specific lexical items and specific phonemes in certain phonetic contexts in which variation would be

¹The authors have been in touch with William Fisher, one of the researchers at Texas Instruments who was involved in the collection of the TIMIT corpus, in an attempt to locate the original documents containing more detailed information about the talkers. Despite his best efforts, however, Dr. Fisher has been unable to locate those documents for us.

expected based on the dialect of the speaker (Fisher et al., 1986; Zue et al., 1990). These two calibration sentences, which were used for all of the acoustic–phonetic analyses reported below, are shown in (1). It should be noted that the use of read speech materials, instead of samples of spontaneous or conversational speech, may have resulted in utterances that do not reflect the true vernacular of the talkers, but instead represent a more formal register of speech (Labov, 1994).

- (1) a. She had your dark suit in greasy wash water all year.
 b. Don't ask me to carry an oily rag like that.

The two TIMIT calibration sentences for each talker were copied into individual sound files that were segmented to include only the sentence material. For the purposes of acoustic analysis, the sound files were leveled to 55 dB using Level16 (Tice & Carrell, 1998).

2.3. Procedure

Eleven acoustic measures were obtained for each of the 66 talkers from the two test sentences. These measures are shown in Table 1. All of the measurements were made using Syntrillium's CoolEdit 96 signal analysis software. The duration measurements were made directly from the spectrogram. Formant frequency measurements were made using the frequency analysis tool in CoolEdit 96, with a 1024 point 50 ms Hamming FFT window. Frequency measurements taken at the "midpoint" were made at the temporal midpoint of the vowel, estimated by visual inspection of the spectrogram. Vowel duration was not considered for any of the vowel measures. Frequency measurements taken at the "onset" were made at the temporal point marking the first third of the vowel. Frequency measurements taken at the "offset" were made at the second to last glottal pulse of the vowel. All frequency measurements were taken at the peak of a glottal pulse.

In order to normalize the frequency measures across the different talkers, the maximum value of the F_2 in the word 'year' was measured for each talker. The location of this local maximum was typically in the first third of the vowel and therefore would not be greatly affected by the following [ɹ]. The motivation for selecting this particular measure was that the maximum F_2 in the vowel /i/ in 'year' should indicate the front-most edge of a given talker's vowel space (Ladefoged, 1993). A comparison of this measure to the F_2 measures of other vowels can be used to determine the relative backness of the other vowels in the talker's space. Because all of the talkers used in this experiment were male, the differences due to vocal tract size were expected to be minimal, but taking relative backness measures instead of absolute backness measures provided a more stable data set.

The eleven acoustic measures were originally selected because they were expected to reveal differences in speech production between the talkers from the six dialect regions. Four of the acoustic measures were obtained from consonants and the remaining seven from vowels. Of the seven vowel measures, three assessed degree of backness and four assessed degree of diphthongization. As is typically the case in variationist research, each vowel was assumed to be phonemically equivalent across all of the 66 talkers (Wells, 1982; Thomas, 2001; Labov et al., in press).

2.3.1. Consonants—Many New England talkers and some Southern talkers are r-less (Kurath & McDavid, 1961; Levine & Crockett, 1971; Labov, 1972). To obtain a measure of r-fulness, the F_3 transition in 'dark' was measured by subtracting F_3 at the offset of the vowel from F_3 at the midpoint of the vowel. In order to have fairly homogenous talkers from each region, all of the New England talkers were r-less, but none of the Southern talkers that were selected were r-less. We therefore predicted that the F_3 transition for the

New England talkers would be smaller than the F_3 transition for the talkers from the other five dialects.

Two consonant alternations were expected to distinguish the South Midland talkers from the other five dialect groups. An alternation between ‘wash’ and ‘warsh’ is found in some South Midland talkers (Kurath & McDavid, 1961; Murray, 1993). This epenthetic [r] has the effect of darkening the preceding vowel. We therefore predicted that the South Midland talkers would have darker vowels in ‘wash’ than talkers from the other dialects. In order to provide some measure of the effects of this alternation on the brightness of the preceding vowel, the midpoint of F_3 in ‘wash’ was measured.

Another lexical alternation that occurs in South Midland as well as Southern speech is the ‘greasy’ ~ ‘greazy’ alternation (Atwood, 1950; Kurath & McDavid, 1961; Wolfram & Schilling-Estes, 1998). We expected that the voiced proportion of the fricative would be greater in the word ‘greasy’ for the Southern and South Midland talkers than for the other talkers. The fricative voicing measure was the proportion of the fricative that was voiced, where voicing was measured categorically as either present or absent. Given the relationship between voicing and duration in consonants of American English, we also predicted that the duration of the fricative would be shorter in ‘greasy’ relative to the length of the entire word for the Southern and South Midland talkers. The fricative duration measure was the ratio of the duration of the fricative to the duration of the entire word.

2.3.2. Vowels—Southern talkers produce more fronted /u/ vowels relative to northern dialect regions (Kurath & McDavid, 1961; Syrdal, 1996; Labov & Ash, 1997). Western talkers also demonstrate a similar trend of fronted /u/ productions (Syrdal, 1996; Hagiwara, 1997). Western and Southern talkers were therefore expected to have fronted /u/’s and, as a result, have smaller relative /u/ backness values than talkers from the other regions. Northern talkers tend to produce more rounded /oo/’s than talkers from the other regions, and this should be reflected acoustically in a lower F_2 (Hillenbrand, Getty, Clark, & Wheeler, 1995; Labov et al., in press). Thus Northern talkers are expected to show a greater relative /oo/ backness value than the other talkers. The relative backness of the /æ/ vowel should be smaller for Northern talkers than for any of the other regions due to the upward and forward movement of /æ/ as part of the Northern Cities vowel shift (Labov et al., 1972; Labov, 1994; Wolfram & Schilling-Estes, 1998). The relative backness of these three vowels was measured in the words ‘suit,’ ‘don’t,’ and ‘rag.’ For each talker, the midpoint of F_2 in ‘suit’ was measured and then subtracted from the maximum F_2 in ‘year’ to obtain a relative backness value of the /u/ vowel. Similarly, the F_2 at the midpoint of ‘don’t’ and ‘rag’ were measured and then subtracted from the maximum F_2 in ‘year’ to obtain relative backness values for the vowels /oo/ and /æ/.

The diphthongization measure for the /oo/ in ‘don’t’ was also expected to separate the Northern talkers from the others. Northern talkers typically show less diphthongization of this vowel (Thomas, 2001; Labov et al., in press). Similarly, Southern talkers were expected to show less diphthongization of the /aɪ/ in ‘like’ and the /oɪ/ in ‘oily,’ because there is a tendency for talkers from this region to produce /aɪ/ as monophthongal /a:/ and /oɪ/ as monophthongal /ɔ:/ (Crane, 1977; Thomas, 2001). Measures of diphthongization were taken by subtracting the offset of F_2 from the midpoint of F_2 in each of the vowels. There is also some evidence in the literature that the /æ/ in ‘rag’ is diphthongized in certain urban regions in the northeast (Kurath & McDavid, 1961; Labov et al., in press). Based on this observation, we predicted that greater diphthongization might be found for this vowel in the speech of New England talkers. The /æ/ diphthong was measured by subtracting the offset of F_2 from the onset of F_2 , in order to magnify any potential differences between dialect groups.

It should be noted here that all of the diphthongization measures rely on at least one formant value measured near the temporal edge of a vowel, such as the onset or the offset. Formant frequency measurements taken near vowel edges are more likely to be affected by the consonantal context than those taken at the midpoint of a vowel. For example, the offglide of the /oo/ in 'don't' is likely to be nasalized due to the following nasal consonant and its second formant may be attenuated as a result of interaction with nasal cavity resonances. The offglides in the /æ/ in 'rag' and the /aɪ/ in 'like' are likely to exhibit raised second formants as a result of the following velar closure. The nucleus of the /æ/ in 'rag' and the offglide of the /oɪ/ in 'oily' may also be affected by the neighboring liquid. In the case of the /æ/, we might expect to find a lower second formant resulting from lip rounding on the [ɹ]. In the case of the /oɪ/, we might also expect to find a lower second formant in transition to the [l]. Given that the phonetic contexts for the vowels remain constant across all of the talkers, however, we still expected to obtain some significant differences in diphthongization between the different regional dialects.

In summary, New England talkers were expected to differ from the other talkers on the measures of r-fulness and /æ/ diphthongization. Northern talkers were expected to differ from the others on the measures of /oo/ backness and diphthongization and /æ/ backness. South Midland talkers were expected to differ from the others on the measure of vowel brightness. Southern and South Midland talkers were expected to differ from the northern and western talkers on the measures of fricative voicing and duration. Southern talkers were expected to differ from the other talkers on the measures of /aɪ/ and /oɪ/ diphthongization. Finally, Southern and Western talkers were expected to differ from the others on the measure of /u/ backness.

While these acoustic-phonetic properties might be found to serve as predictors of dialect affiliation for these talkers, they are not necessarily the only acoustic-phonetic properties, or even the most important properties, that are associated with any given dialect region. In the field of categorization research, a distinction is often made between "characteristic" and "defining" features. Characteristic features are those physical features that are descriptive, but not essential, in describing an object. Defining features, in contrast, are those features that are necessary to the meaning of an object (Smith, 1978). For example, "made of wood" is a characteristic feature of violins since most, but not all, violins are made of wood. On the other hand, "has three sides" is one of the defining features of a triangle because all triangles must have exactly three sides. This distinction is relevant in human categorization research where participants are routinely asked to make judgments about category membership given a variable set of potentially relevant features (Medin, 1989). In the case of dialect categorization, the acoustic measures identified here that are associated with specific dialect affiliations are characteristic, but not necessarily defining, features of a given dialect.

2.4. Results and discussion

As anticipated, the acoustic analysis revealed several consistent differences in speech production between the six dialects considered here. The means and standard deviations for each of the measures for each dialect group are shown in Table 2. A series of one-way ANOVAs was performed to determine which acoustic measures reliably distinguish talkers of different dialects. The r-fulness measure was significant ($F(5, 60) = 3.4, p < 0.01$), as were the fricative voicing measure ($F(5, 60) = 7.2, p < 0.001$), the fricative duration measure ($F(5, 60) = 4.0, p < 0.01$), the /u/ backness measure ($F(5, 60) = 6.6, p < 0.001$), the /oo/ diphthongization measure ($F(5, 60) = 3.8, p < 0.01$), and the /æ/ backness measure ($F(5, 60) = 3.6, p < 0.01$). Means of the remaining five measures, vowel brightness, /oo/ backness, /æ/ diphthongization, /aɪ/ diphthongization, and /oɪ/ diphthongization were not significantly different across the dialect groups.

Post hoc Tukey tests supported several of our initial predictions regarding differences in speech production. In particular, talkers from New England differed significantly from South Midland and Western talkers on mean r-fulness ($p < 0.01$). In the word ‘greasy,’ the mean fricative voicing proportion for Southern talkers differed significantly from New England talkers ($p < 0.01$) and the mean fricative duration for Southern talkers differed significantly from Northern talkers ($p < 0.01$). The mean value of /u/ backness for South Midland, Southern, and Western talkers differed significantly from New England talkers, and /u/ backness was also significantly different between South Midland and Northern talkers (all $p < 0.01$). Degree of /oo/ diphthongization was also significantly different for Northern and Southern talkers ($p < 0.01$). Finally, Northern and New England talkers were significantly different in terms of /æ/ backness ($p < 0.01$).

In order to determine which acoustic–phonetic properties were good predictors of actual dialect affiliation, a step-wise logistic multiple regression² was performed on all 66 talkers for each of the six dialect groups. Dialect affiliation was treated as a dichotomous variable, such that the eleven talkers from each dialect were given a “1” for that dialect and the remaining 55 talkers were given a “0.” Each of the eleven acoustic measures served as a potential predictor variable and was treated as a continuous variable expressed either as a measure of frequency (in Hz) or as a proportion (in the case of the two fricative measures).

Results of the regression analysis are shown in Table 3. An examination of Table 3 reveals that r-fulness and /æ/ backness were the only significant predictor variables for New England talkers ($r^2=0.33$). Given the negative coefficient for r-fulness, these results suggest that New England talkers are r-less and have more /æ/ backing than the other five talker groups. /oo/ and /æ/ diphthongization were significant predictor variables for Northern talkers ($r^2=0.21$). In the case of the /oo/ diphthongization, this result can be interpreted as /oo/ offglide centralization in Northern speech production. For /æ/ diphthongization, the result reflects a monophthongal /æ/. /oo/ and /u/ backness were significant predictor variables for South Midland talkers ($r^2=0.19$). The /u/ backness coefficient is negative and therefore suggests /u/ fronting by South Midland talkers. The /oo/ backness coefficient is positive, however, and suggests /oo/ backing. Voicing of the fricative in ‘greasy’ was a significant predictor of Southern talkers ($r^2=0.21$). There were no significant predictors for either the North Midland or Western talkers.

In summary, the results of the present set of acoustic analyses confirmed that these 66 talkers can be reliably distinguished by dialect based on a small handful of reliable acoustic–phonetic attributes in speech production. The results of the regression analysis revealed seven acoustic–phonetic properties that reliably predict dialect affiliation for these 66 talkers. These findings are consistent with other results reported in the literature on phonological dialect variation (Labov et al., in press; Thomas, 2001). The acoustic–phonetic properties that were found to predict dialect affiliation in this analysis are relatively stable across all of the talkers in a given dialect. However, stability of an acoustic attribute does not necessarily correspond to sufficiency for recognition by listeners. In a study on consonant recognition, Dorman, Studdert-Kennedy, and Raphael (1977) found that while stop bursts are relatively “invariant” across phonetic contexts for a given place of articulation, they are not “sufficient” cues to place of articulation for listeners. In fact, Dorman et al. found that listeners rely on a combination of stop bursts and vowel formant transitions in consonant identification. The following perceptual experiment was therefore designed to investigate how naïve listeners use acoustic attributes to categorize talkers by

²In cases where the value to be predicted (here, actual dialect affiliation of the talker) is a dichotomous variable, a logistic multiple regression is a more appropriate statistical analysis than a linear multiple regression.

dialect, in order to learn more about the relevance of these acoustic–phonetic properties for dialect recognition and categorization.

3. Experiment 2: perceptual categorization task

3.1. Stimulus materials

The two sentences spoken by the 66 male talkers in Experiment 1 were also used in this study. In addition, a third novel sentence was selected from the eight other sentences available on the TIMIT corpus for each talker. A different novel sentence was selected for each of the 66 talkers, so that no sentence would ever be repeated during the course of the experiment. As in Experiment 1, all three sentences for each talker were reproduced in separate sound files that were segmented to include only the sentence material. All of the sound files were leveled to 55 dB using Level16 (Tice & Carrell, 1998).

3.2. Listeners

Twenty-three Indiana University undergraduates served as listeners for this study. All listeners received partial credit for an introductory psychology course for their participation. Data from five of the listeners were removed prior to the final analysis: two were non-native speakers of American English and three performed statistically at chance in all phases of the categorization task.³ The 18 remaining listeners, five males and thirteen females, were all monolingual native speakers of American English with no prior history of hearing or speech disorders.

These 18 listeners were divided post hoc into three listener groups based on their residential history. The seven listeners who had only lived in northern Indiana (north of, and including, Indianapolis) prior to attending school in Bloomington comprised the Northern Indiana group. The five listeners who had lived only in southern Indiana comprised the Southern Indiana group. The remaining six listeners who had all lived out of state for some period of time prior to attending school in Bloomington comprised the Out-of-State group.

3.3. Procedure

The listeners were seated at personal computers equipped with KeyTec Inc. pressure sensitive activation touchscreens (KTMT1315 ProE). The six response categories were displayed as icons on the touchscreens that were depicted as partial maps of the United States, including state boundaries, and were labeled with the name of the dialect region. The six regions are shown in Fig. 2 as they were arranged on the screen. The regions were roughly 2" × 2" in dimension and adequate space was left between the regions to minimize confusions and errors in the response process.

The perceptual experiment was divided into three phases. Prior to beginning the experiment, the maps were displayed on the screen and the listeners were encouraged to familiarize themselves with the six regions. In the first phase, the listeners responded to the first calibration sentence (1a, repeated in 2a) as spoken by each of the 66 talkers one time, presented in random order. On each trial, listeners heard the sentence produced by one of the 66 talkers, presented over headphones (Beyerdynamic DT100) at 70 dB SPL. The listeners were instructed to listen to each sentence carefully and then select the region on the screen that they thought the talker was from. The listeners made their responses by touching the region on the screen. The listeners received no feedback about the accuracy of their

³It is not possible to determine if these three participants were unable to perform the categorization task, despite their best efforts, or if their poor performance was due to a lack of attention to the task itself. Given the homogenous nature of the performance of the remaining participants, the latter explanation seems more likely. For this reason, the data were excluded from the final analysis.

responses. The second phase was identical to the first, except that the listeners responded to the second calibration sentence (1b, repeated in 2b) as spoken by each of the 66 talkers one time, presented in random order. The third phase was identical to the previous two, except that the listeners responded to a novel sentence as spoken by one of the 66 talkers on each trial. Each novel sentence was presented one at a time, in random order, and no sentence was ever presented more than once during the course of the experiment.

- (2) a. She had your dark suit in greasy wash water all year.
- b. Don't ask me to carry an oily rag like that.

3.4. Results and discussion

Overall performance on the six-alternative forced-choice categorization task was poor, but above chance. As shown in Table 4, listeners in the Out-of-State group, Northern Indiana group, and Southern Indiana group performed similarly in terms of proportion correct categorization in all three phases of the experiment. A two-way ANOVA on sentence (first, second, or novel) and listener group (Northern Indiana, Southern Indiana, or Out-of-State) revealed a significant main effect of sentence ($F(2, 15) = 5.1, p = 0.01$). Neither the main effect of listener group nor the sentence-listener group interaction were significant in this analysis. A post hoc Tukey test on sentence revealed that performance on the first calibration sentence and on the novel sentences was significantly better than performance on the second calibration sentence ($p < 0.05$ for both).

Collapsing the data across the three listener groups, the results revealed that the listeners were only able to correctly identify where 33% of the talkers were from on the first calibration sentence, where 28% of the talkers were from on the second calibration sentence, and where 33% of the talkers were from on novel sentences. While overall performance was low, categorization was statistically above chance for all three phases of the experiment. Proportion correct responses are shown in Fig. 3 for each dialect region for each sentence and collapsed across all talkers for each sentence.

3.4.1. Individual listeners—In order to determine if there were any significant effects due to individual listeners, a two-way ANOVA on sentence (first, second, or novel) and individual listener was carried out for each listener group. The main effect of listener and the sentence by listener interaction were not significant for all three groups. Due to the small number of listeners in each group, the main effect of sentence reached significance only for the Out-of-State group ($F(2, 5) = 3.73, p < 0.05$).

3.4.2. Individual talkers—The effects of talker were analyzed further using a series of two-way ANOVAs on sentence (first, second, or novel) and individual talker for each dialect group. The results revealed a significant main effect of individual talker for New England ($F(10, 30) = 5.80, p < 0.001$), North ($F(10, 30) = 2.09, p < 0.05$), North Midland ($F(10, 30) = 2.01, p < 0.05$), and South ($F(10, 30) = 6.37, p < 0.001$). There was also a significant sentence–talker interaction for New England ($F(2, 10) = 3.86, p < 0.001$), North ($F(2, 10) = 1.64, p < 0.05$), North Midland ($F(2, 10) = 2.15, p < 0.01$), and South ($F(2, 10) = 3.17, p < 0.001$). The main effect of sentence was significant only for New England ($F(2, 30) = 20.87, p < 0.001$) and South ($F(2, 30) = 5.38, p < 0.01$). For the New England talkers, post hoc Bonferroni analyses revealed that performance on Sentence 1 was significantly better than performance on Sentence 2 or on novel sentences ($p < 0.001$ and $p < 0.01$, respectively) and performance on the novel sentences was significantly better than performance on Sentence 2 ($p < 0.01$). Post hoc Bonferroni analyses on the Southern talkers revealed that performance on the novel sentences was significantly better than performance on either Sentence 1 or Sentence 2 ($p < 0.05$ for both).

3.4.3. Clustering analysis on perceptual confusions—A stimulus-response confusion matrix was calculated from the responses obtained in the perceptual categorization task for each listener group (Northern Indiana, Southern Indiana, and Out-of-State) for each sentence. An inspection of these confusion matrices based on the perceptual responses suggested that the listeners' difficulty in correctly categorizing a majority of the talkers was not due to random guessing, but in many cases revealed a consistent pattern of perceptual confusions. In order to determine the structure of these responses, the 6×6 confusion matrix for each of the three sentences for each of the three listener groups was submitted to the Similarity Choice Model (SCM; Nosofsky, 1985) to determine similarity and bias parameters between the dialect regions. In the SCM, the probability of response j given stimulus i is a function of the similarity (η) of the stimulus and the response and the bias (b) for that response, as shown in (3). In the present case, the similarity parameters indicate the degree of similarity between each of the dialects, based on the confusion data, whereas the bias parameters indicate the response biases of the listeners.

$$P(R_j|S_i) = \frac{b_j \eta_{ij}}{\sum_k b_k \eta_{ik}} \quad (3)$$

Examination of the bias parameters that resulted from the SCM analysis suggested that the three groups of listeners were not biased to respond with one response alternative more or less often than any of the other response alternatives. For each sentence, the goodness-of-fit of the full SCM for each of the listener groups was compared to the goodness-of-fit of a restricted SCM in which similarity parameters were held constant over all three listener groups. For each sentence, the goodness-of-fit of the full SCM model was significantly better than the goodness-of-fit of the restricted SCM model. This pattern suggests that the three listener groups performed differently in terms of the similarity spaces of the six talker dialects, despite the lack of difference in percent correct categorization performance between the three groups when scored in the conventional manner using only accuracy.

The similarity parameters of the full SCM for each sentence for each listener group were then submitted to an additive clustering scheme,⁴ ADDTREE, to obtain a measure of the perceptual distances between the dialects (Corter, 1995). The resulting trees from the ADDTREE analysis are shown separately for each of the three test sentences in Figs. 4–6. In these figures, perceptual dissimilarity is measured as a function of vertical distance. That is, the dissimilarity between any two dialects is the sum of the lengths of the fewest number of vertical branches connecting the two dialects.

The clustering solutions for the first calibration sentence revealed that the perceptual similarity space for all three listener groups was basically similar in structure, although they varied in terms of perceived distances between dialects. Specifically, all of the listeners appeared to have three broad dialect clusters: New England; South and South Midland; and North, North Midland, and West. The differences between the three listener groups were expressed only in terms of the magnitude of perceptual dissimilarity between any two dialects. By contrast, the differences between the three listener groups for the second calibration sentence and the novel sentences were “structural” in nature because they revealed different dialect clusters. The Out-of-State group had the same three broad clusters

⁴An additive clustering scheme was selected because the initial examination of the confusion matrices indicated that there was a high degree of reciprocity between the six regions. For example, South was most often confused with South Midland and vice versa. This kind of reciprocity is well-modeled by clustering analyses. In addition, other spatial analyses, such as multi-dimensional scaling, were inappropriate for this data given the small size (6×6) of the data matrices.

for the second calibration sentence as for the first. However, the Northern Indiana group displayed a slightly different arrangement of the six dialect regions into three clusters: New England and North; South and South Midland; North Midland and West. The Southern Indiana group, however, had only two main clusters: New England, North, and North Midland; South, South Midland, and West. Finally, for the novel sentences, both Indiana groups again demonstrated the same broad clusters that were found in the first calibration sentence. However, the Out-of-State group revealed the same clusters found in the solution for the Northern Indiana group on the second calibration sentence. Taken together, this analysis revealed variability both between and within listener groups in terms of the perceptual similarity spaces of the dialect regions and provides more detailed insights into the underlying perceptual similarity spaces of these listeners.

3.4.4. Relationship between perceptual categorization and acoustic analysis—

In order to determine which acoustic–phonetic properties were good predictors of the listeners' perceptual categorization data, a series of multiple regression analyses was carried out. For each dialect region, a step-wise linear multiple regression was conducted on all 66 talkers. The acoustic measures were the potential predictor variables and the proportion categorization of each talker into a given dialect region was the value to be predicted. Because the acoustic measures were only obtained from the two calibration sentences, the categorization scores for each talker used in this analysis were based only on the two calibration sentences. All of the data were scored on a continuous scale, either as a proportion or in frequency (Hz).

The results of the regression analyses are shown in Table 5. R-fulness, /æ/ backness, /oo/ diphthongization, and vowel brightness in 'wash' were significant predictor variables for categorization as New England ($r^2=0.39$). As in the previous regression analysis, the negative coefficients for the r-fulness and /oo/ diphthongization variables suggest that r-lessness and a centralized /oo/ offglide are properties listeners attend to in categorizing talkers as New Englanders. The centralization of the /oo/ offglide is also found in the categorization of Northern talkers where /oo/ diphthongization and /u/ backness were significant predictor variables for categorization as Northern ($r^2=0.27$). The only significant predictor of North Midland categorization was /oi/ diphthongization ($r^2=0.31$). /u/ backness, vowel brightness in 'wash,' and fricative voicing in 'greasy' were significant predictor variables of categorization as South Midland ($r^2=0.38$). The negative coefficients for /u/ backness and vowel brightness indicate that /u/ fronting and vowel darkness (i.e., [waɪ ʃ] for [waʃ]) are used by listeners as clues to South Midland dialect affiliation. Fronting of /u/ was also a predictor of categorization as Southern, as revealed by the negative coefficient for the significant /u/ backness variable. The other significant predictor variables for Southern categorization were /oi/ diphthongization, /oo/ diphthongization and backness, and /æ/ diphthongization ($r^2=0.49$). Interpretation of the signs of the coefficients reveals that /oi/ monophthongization, peripheral /oo/ offglides, /oo/ backness, and /æ/ diphthongization were significant acoustic properties for Southern categorization. Finally, /oi/ diphthongization was the only significant predictor of categorization as Western ($r^2=0.16$). The results of the regression analysis suggest that listeners relied on a small set of acoustic–phonetic properties in making their dialect categorization judgments of the talkers used in this study.

The results of the regression analysis follow fairly well from the predictions in Experiment 1. It appears that the naïve listeners were attending to relevant properties of the acoustic signal in making their categorization judgments. In particular, the listeners seemed to be relying on both positive attributes of a given dialect (such as New England r-lessness or Northern /oo/ offglide centralization) as well as negative attributes of a given dialect (such as Northern /u/ backness, which is the equivalent of "not Southern /u/ fronting"). In the case of the Western and North Midland categorization, the listeners were relying on /oi/

diphthongization. Given that the results of the regression analysis in Experiment 1 failed to reveal any acoustic–phonetic properties associated with dialect affiliation for these two groups, the listeners were quite limited in appropriate cues to dialect categorization for the North Midland and Western talkers. It seems that the listeners were simply trying to associate at least one distinctive acoustic property with these two regions and this particular set of listeners appeared to use /oɪ/ diphthongization for that purpose.

4. General discussion

The acoustic analyses performed in the first experiment confirmed that the talkers selected from each dialect region reliably produced phonological differences that can be measured acoustically. As expected, r-lessness was a good predictor of New England talkers, centralized /oo/ offglides were good predictors of Northern talkers, /u/ fronting was a good predictor of South Midland talkers, and the presence of a voiced fricative in ‘greasy’ was a good predictor of Southern talkers. Unexpectedly, /æ/ backness turned out to be a good predictor of New England talkers while /æ/ monophthongization was a good predictor of Northern talkers. The high degree of /æ/ backness found in the New England talkers might be indicative of the predicted diphthongization in which the offglide is realized as a centralized vowel such as [ə], in which case the high degree of backness would reflect this centralization. Alternatively, the /æ/ backing could reflect true backing of the vowel to something like [ɑ] or [ɒ], although this variant seems to be more lexically restricted (Kurath & McDavid, 1961). Given the small speech sample available in the present study, discriminating between these two interpretations is not feasible. In contrast, the small amount of diphthongization of /æ/ in the North might reflect its raising due to the Northern Cities shift. In particular, the effect of coarticulation with the following [g] in ‘rag’ may be reduced for Northern Cities speakers who have fronted /æ/’s, resulting in higher second formants in the beginning of the vowel than speakers without a raised and fronted /æ/. Finally, /oo/ backness was an unexpected predictor of South Midland talkers, given that both its nucleus and its glide are typically fronted in Southern and South Midland speech as part of the Southern shift. However, the fronting of /oo/ is a change in progress and follows the fronting of /u/ (Thomas, 2001). It is therefore possible that the fronting of /oo/ is not reflected in the talkers used in the present study because they had not yet adopted a fronted /oo/ at the time the original TIMIT recordings were made in the late 1980s.

Several of the acoustic–phonetic measures that were expected to reveal differences between the dialect groups did not reliably predict the dialect affiliation of any of the six groups of talkers. Specifically, we expected vowel brightness in ‘wash’ to distinguish the South Midland from the other five dialects. However, this measure was not a strong predictor of South Midland dialect affiliation. One reason for this result may be that this measure is based on an absolute measure of F_3 , which means that it was not normalized across speakers for vocal tract size, unlike the measures involving F_2 which were normalized against the F_2 of ‘year’ to account for talker differences. Measures of vowel brightness are also potentially problematic because the vowel itself is associated with varying degrees of rounding in different dialects, which has an acoustic impact on F_3 .

The measure for the diphthong /aɪ/ was predicted to distinguish the Southern talkers from the others. However, it turned out to be a weak predictor of Southern dialect affiliation. The measure of degree of diphthongization of /aɪ/ was also somewhat problematic in this analysis because it was taken from the word ‘like.’ The monophthongization process is weaker before voiceless consonants than voiced in some varieties of American English, so we might not expect to see as much monophthongization in the word ‘like,’ even in varieties with some /aɪ/ monophthongization (Thomas, 2000). Second, as mentioned above, a following velar context generally results in an upward offglide of the preceding vowel

(Ladefoged, 1993). This upward offglide may have concealed the expected monophthongization of /aɪ/ in the Southern talkers.

While the fricative voicing in 'greasy' was a significant predictor of Southern talkers, fricative duration was not a significant predictor of either the Southern or the South Midland talkers. Given that consonant voicing and duration are strongly linked in English, it is interesting that this measure did not turn out to be significantly related to the Southern talkers. Finally, the measure of /oɪ/ diphthongization was also expected to distinguish the Southern talkers from the others, but was not a significant predictor of Southern dialect affiliation for these talkers. This particular variable is less common for white speakers than some of the other variables examined in this study, which may account for the non-significant result (Thomas, 2001).

In addition to potential explanations for the lack of expected results due to the nature of some of the acoustic measurements, the specific speech samples that were used in this study must be considered. First, the present analysis was based on a relatively small number of acoustic-phonetic measurements, due to the limitations of the TIMIT corpus. With additional speech materials, it would be possible to obtain a more powerful assessment of the acoustic-phonetic properties of the speech signal. Second, it is possible that the regions used to define the talkers in this study were not the most accurate categorization of these talkers. For example, several dialect researchers have suggested that the Midland areas should be considered as one single homogeneous region while others suggest that it should not be treated as a unique region at all (Carver, 1987; Johnson, 1991; Davis & Houck, 1992; Labov et al., in press). There is also some controversy about the extent of variation, or lack thereof, in the vast geographical area contained within the Western dialect region (Labov et al., in press).

It is also unclear what criteria were used to assign the dialect labels to the talkers in the compilation of the TIMIT corpus. Because the TIMIT corpus was designed originally for speech recognition research, the goal was simply to collect a speech corpus that contained large amounts of variation (Fisher et al., 1986; Zue et al., 1990). The representativeness of the talkers with respect to their dialect region was therefore not an important issue at the time the corpus was collected. Questions such as where a particular talker lived (e.g., city and state), how long he had lived in the dialect region he represented, and whether or not his parents were also from that dialect region were likely not relevant in indicating dialect, and such information is not provided with the TIMIT corpus. There is reason to believe, however, that such detailed characteristics of the talker may be quite important if one wants to make broader generalizations about the effects of dialect variation on perceptual categorization as in the present study.

Finally, the individual differences and variability observed between the talkers in a given dialect region may also provide an explanation for some of the unexpected results. The standard deviations of the means shown in Table 2 reveal substantial variation between the talkers within any given dialect group. It is very likely that some of the talkers in a given dialect are better representatives of their region than others. Indeed, the analyses of the perceptual categorization results on individual talkers revealed that some talkers were more easily categorized by listeners than others (and see Bradlow, Torretta, & Pisoni, 1996). In particular, main effects of talker were observed for New England, North, North Midland, and South. One property that three of these regions have in common is the presence of several stereotyped features, such as New England /æ/ diphthongization, Northern /oʊ/ rounding, and Southern /oɪ/ monophthongization. While none of these phonetic properties was a strong predictor of talkers from that region, it is possible that some of the talkers

exhibited these properties and those talkers were better identified than talkers from the same region who did not share those properties.

The main effect of talker for North Midland is initially more puzzling, given the lack of distinctive acoustic–phonetic features of North Midland speakers. However, since one-third of the listeners in the perception study were from northern Indiana, we would expect these listeners to be highly familiar with the North Midland dialect and perhaps be speakers of this dialect themselves. In addition, the remaining two-thirds of the listeners attended school in Bloomington, Indiana and we would also expect them to be familiar with the North Midland dialect given the population of students at Indiana University. Perhaps the main effect of individual talker for the North Midland dialect group simply reflects this familiarity on the part of the listeners with this dialect and the variability that is found within it (Black & Tolhurst, 1955; Clopper & Pisoni, submitted).

The results of the talker analysis also revealed significant talker–sentence interactions for New England, North, North Midland, and South. These interactions suggest that the listeners' success in categorizing the talkers was based on their use of available cues in the sentence. For example, listeners were attending to the centralized offglide in /oo/ in the second sentence in order to identify Northern talkers and they were attending to /u/ backness in the first sentence to identify Northern talkers. However, only the centralized /oo/ offglide is actually a reliable predictor of Northern talkers, which means that talkers exhibiting that property might be better identified for the second sentence than others, whereas talkers exhibiting /u/ backness in the first sentence might be better identified, regardless of the fact that /u/ backness was not a strong predictor of this set of Northern talkers.

The role of the availability of reliable acoustic–phonetic cues in speech is most apparent for categorization performance on the New England talkers for whom there was a significant effect of sentence. Categorization performance of the New England talkers was significantly better on the first sentence than on the second. Recall that the cue to r-lessness, which was both a good predictor of talkers from New England and a good predictor of categorization as New England, was robustly present in the first calibration sentence in the word 'dark,' but was not in the second calibration sentence. The difference in categorization performance between the two calibration sentences for these talkers may be attributed to this difference in availability of this important acoustic–phonetic cue.

The overall results of the perceptual categorization experiment revealed that the listeners used in this study were able to correctly categorize the talkers by dialect with only about 30% accuracy. These results are surprisingly quite similar to those obtained by Williams et al. (1999) in their study of the categorization of Welsh English dialects. Both our results and those of Williams et al. are somewhat lower than those obtained by Van Bezooijen and Gooskens (1999) for the perceptual categorization of British and Dutch dialects, whose listeners exhibited 52% and 35% accuracy, respectively. While Clopper and Pisoni (2002) found that short-term exposure in the laboratory to dialect variation led to slightly improved performance on the perceptual categorization of unfamiliar talkers, the results reported here and by Williams et al. and Van Bezooijen and her colleagues suggest that the task of categorizing an unfamiliar talker by his or her dialect based on a short sample of speech is quite difficult for a naïve untrained listener.

There are several possible explanations for why this task appears to be so difficult for naïve listeners. First, in normal listening situations, listeners are rarely, if ever, required to make explicit judgments about where an unfamiliar talker is from based on exposure to only a single sentence. If listeners need to make this kind of judgment at all, they are able to base their decisions on more than a single utterance and can use other sources of information

about the talker, such as his style of dress or the topic of conversation, to make their judgments. Second, the labels and categories that were given to the listeners as response alternatives in the recent dialect categorization studies including the present study, might not have corresponded to the categories that the listeners typically used to identify where talkers are from. If the listeners were accustomed to using different categories or a different number of labels, the task might become artificially more difficult due merely to the mismatch between the participants' prior perceptual categories and the response alternatives provided by the experimenters. As mentioned above with respect to Preston's (1986, 1989, 1993) perceptual dialectology research, listeners' knowledge of United States geography may also be partly to blame for their poor performance. Although three of the partial maps included Texas, two included Indiana (where all the participants attend school), and the maps were all labeled with somewhat familiar names, it is still possible that participants were limited by incomplete knowledge of geography. Finally, as in the acoustic analysis experiment, the use of read speech materials may have led the talkers to use more standard forms of speech, making dialect categorization more difficult for the listeners.

The clustering analysis of the results of the categorization task in the second experiment demonstrated that naïve listeners might represent dialect variation in terms of broader categories than those presented as response alternatives in the perceptual study which were taken from the TIMIT corpus and based on outdated sociolinguistic research. Specifically, the listeners in this experiment grouped the 66 talkers into the same three broad dialect categories in six of the nine solutions (two of the three sentence conditions per listener group): New England; South and South Midland; and North, North Midland, and West. Of the remaining three solutions, two also revealed three broad clusters with a slightly different composition: New England and North; South and South Midland; and North Midland and West. Despite this variation between and within listener groups, however, this clustering analysis suggests that the responses were not random, but that listeners were sensitive to perceptual similarities between the six dialects and that they were perhaps basing their categorization judgments on three broad categories, as opposed to the six smaller dialect regions they were given as response alternatives. In particular, the results suggest that the listeners in this study were making most of their judgments based on three clusters: New England (and sometimes North); South and South Midland; North, North Midland, and West.

The division of the United States into three broad dialect categories is consistent with previous research on dialect variation conducted by sociolinguists. For example, Preston (1993) found that listeners could group his nine talkers into roughly a Northern group and Southern group, suggesting that when the east–west dimension is removed, listeners tend to base their identification on a major north–south boundary. Many years ago, Krapp (1925) referred to three major varieties of American English in his discussion of phonological variation: Eastern (roughly New England), Southern, and Western or “General” American. Finally, the extensive research done by Labov and his colleagues on the description of dialect variation of the United States has led him to propose three major dialects of American English that capture much of the phonological variation in North America: Northern, Southern, and Western (Labov, 1998). It is not surprising, then, that in a perceptual task where lexical variation is held constant due to the nature of the read speech materials listeners would rely on three major dialects based on phonological variation, even if they could not make more precise judgments about the six regions provided as response alternatives in this study.

In addition to the novel structural results obtained from the clustering analysis, several notable aspects of the underlying perceptual similarity spaces were also revealed in these analyses. In eight of the nine clustering solutions, the Southern dialect was more dissimilar

from the other dialects than the South Midland was. It can be seen from the clustering results that in all cases, the South and South Midland are clustered fairly early in the tree, but the vertical line connecting the South to this South–South Midland node is typically longer than the vertical line connecting the South Midland to this node. This pattern suggests that the listeners were sensitive to differences between the Southern and South Midland dialects. In particular, the greater distance between the South and the other dialects than the South Midland suggests that the South is perceptually more distinct than the South Midland. The relationship between physical distance in the clustering solution and perceptual distinctiveness can also be seen in the typically long branch connecting New England to the rest of the tree. Like the South, listeners also perceived New England as being more distinctly dissimilar with respect to the other dialects.

No significant effects of either listener group (defined by a post hoc division into one of three groups based on residential history) or individual listeners within a given group in terms of categorization accuracy were found. The lack of group differences can likely be attributed to the small number of listeners in each subgroup and the fact that the listeners were divided post hoc into the three groups. However, the SCM analyses on the confusion matrices produced by each group did reveal inter-group differences in terms of their underlying perceptual dialect similarity spaces. Further study of the effects of residential history on performance in perceptual tasks is clearly warranted. Bayard, Weatherall, Gallois, and Pittam (2001) have found differences in their work on evaluative responses to New Zealand, Australian, American, and British English by listeners from different countries. And, in a more recent study, Clopper and Pisoni (submitted) have also found listener group differences based on residential history in the same dialect categorization task described in Experiment 2 above. The lack of individual differences within each group in this study suggests that some listeners are not inherently better at this task than others, and that this task is equally difficult for all individuals. Again, however, individual listener differences may be an important issue to pursue in future perceptual studies of dialect variation (Mason, 1946).

The regression analysis in the first experiment revealed that seven acoustic–phonetic properties were found to be good predictors of actual dialect affiliation of the talkers. That is, seven good acoustic properties were available to the listeners in the perceptual categorization task. Despite the overall difficulty of this task, however, the listeners demonstrated reliable use of 16 acoustic cues in categorizing the talkers. Of the seven available cues and the 16 attended cues, four were both available to and used by the listeners: New England r-lessness, New England /æ/ backness, Northern /oo/ offglide centralization, and South Midland /u/ fronting.

The fact that New England r-lessness and the Northern /oo/ offglide were both available and attended to is not too surprising. R-lessness is a stereotypical feature of Boston English that is highlighted by the saying, ‘Pahk the cah in Hahvahd yahd’ [pɑ:k ðə kɑ: ʔɪn hɑ:vəd jɑ:d]. Similarly, the distinctiveness of the Northern /oo/ is stereotyped in the media, in such films as *Fargo*. The sensitivity of the listeners to New England /æ/ backing and South Midland /u/ fronting was more surprising. As discussed above, the /æ/ backing found in the New England talkers was an unpredicted result, but it likely reflects either true backing, similar to that found in some varieties of British English, or diphthongization. The fact that New England speakers are often characterized as maintaining British forms might account for the listeners’ use of the similarity between a backed /æ/ of one of these talkers and a British /æ/ as a cue to New England. In addition, /æ/ diphthongization is often associated stereotypically with the New York area, despite its use in other parts of the country.

While /u/ fronting is not something that is as easily identified as being part of an American English speaker's set of dialect stereotypes, it was used correctly by the listeners in this study to identify talkers from the South Midland area. Perhaps the listeners in this study were more sensitive to that particular cue than others might be due to their location in southern Indiana, where the local speech is probably best characterized as South Midland. Because /u/ fronting is common in the South and West in addition to the South Midland, it is perhaps the location of the listeners themselves that best explains their use of this particular cue. As Preston (1986, 1993) has demonstrated in his perceptual dialectology work, where the participant lives has an impact on how she or he constructs mental maps of dialect variation and identifies talkers.

Finally, we should emphasize here that the listeners in this study were not given any training, experience, or feedback with the talkers before they participated in the perceptual categorization task. The results of this study reflect the ability of the listeners to use whatever knowledge they had acquired as native speakers of American English before they came into the laboratory to identify where the listeners were from. In a follow-up questionnaire which was completed immediately after the perceptual categorization task, the listeners were asked to identify properties that they had listened for in trying to categorize the talkers. The survey responses included such comments as specific sounds ('o' or 'a') or words ('greasy' or 'wash'), as well as specific accents (New York or Boston, Southern 'drawl' or 'twang'). Although these informal responses are subject to the same kinds of criticisms as attitude judgment research, particularly with regard to the role of stereotypes in perception, they reveal that listeners do have some explicit knowledge and awareness of some of the properties that characterize regional dialects of American English, without having undergone any explicit training with the individual talkers used in this study or exposure to regional dialect variation in general.

5. Conclusions

The results of the first experiment using acoustic measurement techniques provide further evidence that phonological differences do exist between regional dialects of American English and that the dialect affiliation of the talkers can be predicted to some extent by well-defined acoustic-phonetic differences in speech, even short samples of read speech, such as sentences. The results of the second perceptual categorization experiment replicated previous work by Williams et al. (1999) which showed that listeners can only categorize unfamiliar talkers by dialect with about 30% accuracy. The perceptual findings also provide additional support for Preston's (1993) conclusion that naïve listeners do not necessarily categorize talkers accurately by dialect region, but they are able to make reliable distinctions between dialect groups using broad perceptual categories. In particular, the results of the clustering analysis suggested that the naïve listeners in Experiment 2 were able to reliably identify talkers from New England, South, and West regions, which roughly correspond to the three major dialects of American English described by Krapp (1925) and Labov (1998). While there were no differences between the three listener groups based on categorization performance, residential history did appear to play a role in shaping the underlying perceptual similarity spaces of the dialects. Further research is needed to investigate the contribution of residential history in dialect categorization. Finally, all of the listeners in this study appeared to rely on a relatively small set of acoustic-phonetic properties in making their perceptual categorization judgments. Taken together, the results of these two experiments confirm the intuition that naïve listeners have an explicit awareness of several highly distinctive and discriminable phonological differences between dialects of American English and can use this knowledge of phonological variation to reliably categorize talkers by dialect region, without any specific training or feedback.

Acknowledgments

This work was supported by the NIH-NIDCD R01 research grant DC00111 and the NIH-NIDCD T32 training grant DC00012 to Indiana University. We would like to thank Caitlin Dillon for her assistance in selecting the talkers for this project, Luis Hernandez for his technical advice and support, Kenneth de Jong for suggesting some of the measures for the first experiment, Robert Nosofsky for his help in conducting the clustering analyses in the second experiment, and James Harnsberger, Allyson Carter, and the reviewers for their comments and valuable suggestions on earlier versions of this paper.

References

- Atwood EB. Grease and greasy: A study of geographical variation. *Studies in English* (University of Texas). 1950; 29:249–260.
- Bayard D, Weatherall A, Gallois C, Pittam J. Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics*. 2001; 5:22–49.
- Black JW, Tolhurst GC. The relative intelligibility of language groups. *Quarterly Journal of Speech*. 1955; 41:57–60.
- Bourhis RY, Giles H, Lambert WE. Social consequences of accommodating one's style of speech: A cross-national investigation. *International Journal of the Sociology of Language*. 1975; 6:55–72.
- Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*. 1996; 20:255–272.
- Bush CN. Some acoustic parameters of speech and their relationships to the perception of dialect differences. *TESOL Quarterly*. 1967; 1(3):20–30.
- Byrd D. Sex, dialects, and reduction. *Proceedings of the International Conference on Spoken Language Processing—ICSLP'92 Proceedings*. 1992; 92:827–830.
- Byrd D. Relations of sex and dialect to reduction. *Speech Communication*. 1994; 15:39–54.
- Carver, CM. *American regional dialects: A word geography*. Ann Arbor, MI: University of Michigan Press; 1987.
- Clopper, CG.; Pisoni, DB. Effects of talker variability on perceptual learning using a dialect categorization task. Poster presented at *New Ways of Analyzing Variation 31*; Stanford, CA. October 10–13; 2002.
- Clopper CG, Pisoni DB. Homebodies and army brats: some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*. submitted.
- Cortier JE. ADDTREE/P Program for Fitting Additive Trees. 1995
- Crane, LB. The social stratification of /ai/ in Tuscaloosa, Alabama. In: Shores, DL.; Hines, CP., editors. *Papers in language variation*. Tuscaloosa, AL: University of Alabama Press; 1977. p. 180–200.
- Davis LM, Houck CL. Is there a Midland dialect?—again. *American Speech*. 1992; 67:61–70.
- Docherty, GJ.; Foulkes, P. Derby and Newcastle: instrumental phonetics and variationist studies. In: Foulkes, P.; Docherty, GJ., editors. *Urban voices: accent studies in the British Isles*. London: Edward Arnold; 1999. p. 47–71.
- Dorman MF, Studdert-Kennedy M, Raphael LJ. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*. 1977; 22:109–122.
- Fisher WM, Doddington GR, Goudie-Marshall KM. The DARPA speech recognition research database: Specifications and status. *Proceedings of the DARPA Speech Recognition Workshop*. 1986:93–99.
- Giles H. Evaluative reactions to accents. *Educational Review*. 1970; 22:211–227.
- Giles H, Bourhis RY. Dialect perception revisited. *Quarterly Journal of Speech*. 1973; 59:337–342.
- Hagiwara R. Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*. 1997; 102:655–658.
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*. 1995; 97:3099–3111. [PubMed: 7759650]

- Inoue, F. Subjective dialect division in Great Britain. In: Preston, DR., editor. *Handbook of perceptual dialectology*. Philadelphia: John Benjamins; 1999. p. 161-176.
- Johnson E. Yet again: The midland dialect. *American Speech*. 1991; 69:419–430.
- Keating P, Blankenship B, Byrd D, Flemming E, Todaka Y. Phonetic analyses of the TIMIT corpus of American English. *Proceedings of the International Conference on Spoken Language Processing —ICSLP'92 Proceedings*. 1992; 92:823–826.
- Keating PA, Byrd D, Flemming E, Todaka Y. Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication*. 1994; 14:131–142.
- Krapp, GP. *The English language in America*. New York: Frederick Ungar; 1925.
- Kurath, H., editor. *The linguistic atlas of New England*. Providence: Brown University Press; 1939.
- Kurath, H.; McDavid, RI. *The pronunciation of English in the Atlantic States*. Ann Arbor, MI: University of Michigan Press; 1961.
- Labov, W. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press; 1972. The social stratification of (r) in New York City department stores; p. 43-69.
- Labov, W. *Principles of linguistic change: Internal factors*. Cambridge, MA: Blackwell; 1994.
- Labov, W. The three dialects of English. In: Linn, MD., editor. *Handbook of dialects and language variation*. San Diego: Academic Press; 1998. p. 39-81.
- Labov, W.; Ash, S. Understanding Birmingham. In: Bernstein, C.; Nunnally, T.; Sabino, R., editors. *Language variety in the south revisited*. Tuscaloosa, AL: University of Alabama Press; 1997. p. 508-573.
- Labov, W.; Ash, S.; Boberg, C. *Atlas of North American English*. Berlin: Mouton de Gruyter; in press
- Labov, W.; Yaeger, M.; Steiner, R. *A quantitative study of sound change in progress*. Philadelphia: US Regional Survey; 1972.
- Ladefoged, P. *A course in phonetics: third edition*. Fort Worth, TX: Harcourt Brace; 1993.
- Lambert W, Hodgson ER, Gardner RC, Fillenbaum S. Evaluation reactions to spoken languages. *Journal of Abnormal and Social Psychology*. 1960; 60:44–51. [PubMed: 14413611]
- Levine, L.; Crockett, H. Speech variation in a Piedmont Community. In: Williamson, J.; Burke, V., editors. *A various language: perspectives on American dialects*. New York: Holt, Rinehart, & Winston; 1971. p. 437-460.
- Mason HM. Understandability of speech in noise as affected by region of origin of speaker and listener. *Speech Monographs*. 1946; 13(2):54–58.
- Medin DL. Concepts and conceptual structure. *American Psychologist*. 1989; 44:1469–1481. [PubMed: 2690699]
- Murray, TE. The language of St. Louis, Missouri: Dialect mixture in the urban midwest. In: Frazer, TC., editor. *“Heartland” English: variation and transition in the American Midwest*. Tuscaloosa, AL: University of Alabama Press; 1993. p. 125-136.
- Niedzielski N. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*. 1999; 18:62–85.
- Nosofsky R. Overall similarity and the identification of separable-dimension stimuli: A choice-model analysis. *Perception & Psychophysics*. 1985; 38:415–432. [PubMed: 3831920]
- Preston D. Five visions of America. *Language in Society*. 1986; 15:221–240.
- Preston, D. *Perceptual dialectology: Nonlinguists’ views of areal linguistics*. Providence, RI: Foris; 1989.
- Preston, D. Folk dialectology. In: Preston, D., editor. *American dialect research*. Philadelphia: John Benjamins; 1993. p. 333-378.
- Preston, D. The social interface in the perception and production of Japanese vowel devoicing: it’s not just your brain that’s connected to your ear. Paper presented at the 9th biennial Rice University Symposium on Linguistics: Speech Perception in Context; Houston, TX. March 13–16; 2002.
- Purnell T, Idsardi W, Baugh J. Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*. 1999; 18:10–30.
- Ryan EB, Carranza M. Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. *Journal of Personality and Social Psychology*. 1975; 31:855–863.

- Smith, EE. Theories of semantic memory. In: Estes, WK., editor. Handbook of learning and cognitive processes. Vol. 6. Hillsdale, NJ: Erlbaum; 1978. p. 1-56.
- Syrdal AK. Acoustic variability in spontaneous conversational speech of American English talkers. ICSLP 96 Proceedings. 1996; 96:438-441.
- Thomas ER. Spectral differences in /ai/ offsets conditioned by voicing of the following consonant. Journal of Phonetics. 2000; 28:1-25.
- Thomas, ER. An acoustic analysis of vowel variation in New World English. Durham, NC: Duke University Press; 2001.
- Tice, R.; Carrell, T. Level16 (version 2.0.3). Lincoln, NE: University of Nebraska; 1998.
- Van Bezooijen R, Gooskens C. Identification of language varieties: The contribution of different linguistic levels. Journal of Language and Social Psychology. 1999; 18:31-48.
- Van Bezooijen R, Ytsma J. Accents of Dutch: Personality impression, divergence, and identifiability. Belgian Journal of Linguistics. 1999; 13:105-129.
- Wales, K. North and south: a linguistic divide? Inaugural professorial lecture. University of Leeds; 1999 June 10. Retrieved from the World Wide Web December 16, 2002 (<http://www.leeds.ac.uk/reporter/439/kwales.htm>)
- Wells, JC. Accents of English I: An introduction. Cambridge: Cambridge University Press; 1982.
- Williams, A.; Garrett, P.; Coupland, N. Dialect recognition. In: Preston, DR., editor. Handbook of perceptual dialectology. Philadelphia: John Benjamins; 1999. p. 345-358.
- Williams, F. Explorations of the linguistic attitudes of teachers. Rowley, MA: Newbury House; 1976.
- Wolfram, W.; Schilling-Estes, N. American English. Malden, MA: Blackwell; 1998.
- Zue V, Seneff S, Glass J. Speech database development at MIT: TIMIT and beyond. Speech Communication. 1990; 9:351-356.

Fig. 1. Northern Cities vowel shift (left) and Southern Vowel Shift (right). Adapted from Wolfram and Schilling-Estes (1998, pp. 138–139).

Fig. 2.
The six response alternatives in the categorization task.

Fig. 3.

Percent correct categorization of dialect affiliation of the talkers for each sentence, collapsed across the three listener groups. Percent correct categorization collapsed across all six talker groups is shown on the right. Chance performance (17%) is shown with a dashed line, performance statistically above chance (25%) is shown with a solid line.

Fig. 4.
Clustering solution for the first calibration sentence for each of the three listener groups, Out-of-State, Northern Indiana, and Southern Indiana.

Fig. 5. Clustering solution for the second calibration sentence for each of the three listener groups, Out-of-State, Northern Indiana, and Southern Indiana.

Fig. 6. Clustering solution for the novel sentences for each of the three listener groups, Out-of-State, Northern Indiana, and Southern Indiana.

Table 1

Acoustic measures selected for comparison between dialect groups

Word	Segment	Measurement	Acoustic-phonetic property
dark	/a/	F ₃ midpoint–F ₃ offset	r-fulness
wash	/a/	F ₃ midpoint	vowel brightness
greasy	/s/	proportion of fricative that is voiced	fricative voicing
		ratio of fricative duration to word duration	fricative duration
suit	/u/	maximum F ₂ in 'year'–F ₂ midpoint	/u/ backness
don't	/oo/	maximum F ₂ in 'year'–F ₂ midpoint	/oo/ backness
		F ₂ midpoint–F ₂ offset	/oo/ diphthongization
rag	/æ/	maximum F ₂ in 'year'–F ₂ midpoint	/æ/ backness
		F ₂ offset–F ₂ onset	/æ/ diphthongization
like	/aɪ/	F ₂ offset–F ₂ midpoint	/aɪ/ diphthongization
oily	/oɪ/	F ₂ offset–F ₂ midpoint	/oɪ/ diphthongization

Table 2

Means for the acoustic measurements for each of the six talker groups. The standard deviation for each measurement is shown in parentheses after the mean

	New England	North	North Midland	South Midland	South	West
r-fulness (Δ Hz)	262 (92)	409 (124)	358 (149)	462 (155)	422 (121)	451 (145)
vowel brightness (Hz)	2373 (196)	2302 (98)	2330 (136)	2133 (351)	2203 (149)	2179 (230)
fricative voicing (%)	0.07 (0.10)	0.05 (0.06)	0.02 (0.06)	0.27 (0.42)	0.57 (0.49)	0.03 (0.04)
fricative duration (%)	0.33 (0.04)	0.36 (0.03)	0.36 (0.03)	0.34 (0.07)	0.29 (0.07)	0.35 (0.02)
/u/ backness (Hz)	609 (218)	557 (198)	496 (156)	293 (152)	337 (122)	334 (166)
/oo/ backness (Hz)	1004 (150)	1105 (203)	991 (149)	1038 (150)	1012 (129)	939 (203)
/oo/ diphthong (Δ Hz)	-71 (78)	-148 (108)	-40 (130)	22 (167)	37 (113)	-41 (58)
/æ/ backness (Hz)	601 (151)	399 (137)	440 (124)	425 (103)	494 (69)	491 (154)
/æ/ diphthong (Δ Hz)	256 (128)	177 (46)	255 (120)	280 (175)	223 (67)	233 (107)
/aɪ/ diphthong (Δ Hz)	452 (147)	418 (149)	402 (164)	278 (134)	331 (143)	350 (190)
/oi/ diphthong (Δ Hz)	301 (200)	384 (184)	434 (188)	250 (225)	226 (165)	445 (197)

Table 3

Results of the logistic multiple regression analysis on acoustic-phonetic properties and talker dialect affiliation. For each of the dialect groups, the significant acoustic measures are shown with their regression coefficients and the overall r^2 showing model fit

	Significant variables	Regression coefficients	Overall r^2
New England	r-fulness	−0.01	0.33
	/æ/ backness	0.02	
North	/oo/ diphthong	−0.01	0.21
	/æ/ diphthong	−0.01	
North Midland	n/a		
South Midland	/u/ backness	−0.01	0.19
	/oo/ backness	0.01	
South	fricative voicing	3.4	0.21
West	n/a		

Table 4

Proportion correct categorization scores for the three listener groups (Northern Indiana, Southern Indiana, and Out-of-State) in the three phases of the experiment (Sentence 1, Sentence 2, and Novel sentences)

	Sentence 1	Sentence 2	Novel sentences
Northern Indiana ($N = 7$)	0.32	0.28	0.33
Southern Indiana ($N = 5$)	0.35	0.28	0.29
Out-of-State ($N = 6$)	0.34	0.29	0.37

Table 5

Results of the linear multiple regression analysis on the acoustic-phonetic properties and perceptual categorization responses. For each of the dialect groups, the significant acoustic measures are shown with their regression coefficients and the overall r^2 showing model fit

	Significant variables	Regression coefficients	Overall r^2
New England	r-fulness	−0.36	0.39
	/æ/ backness	0.34	
	/oo/ diphthong	−0.22	
	vowel brightness	0.21	
North	/oo/ diphthong	−0.38	0.27
	/u/ backness	0.29	
North Midland	/oi/ diphthong	0.56	0.31
South Midland	/u/ backness	−0.26	0.38
	vowel brightness	−0.34	
	fricative voicing	0.33	
South	/oi/ diphthong	−0.39	0.49
	/oo/ diphthong	0.33	
	/u/ backness	−0.33	
	/oo/ backness	0.31	
	/æ/ diphthong	0.20	
West	/oi/ diphthong	0.40	0.16