Published in final edited form as:

J Phon. 2007 July ; 35(3): 421–438.

Free classification of regional dialects of American English

Cynthia G. Clopper^{*} and David B. Pisoni

Indiana University, Bloomington, IN 47405, USA

Abstract

Recent studies have found that naïve listeners perform poorly in forced-choice dialect categorization tasks. However, the listeners' error patterns in these tasks reveal systematic confusions between phonologically similar dialects. In the present study, a free classification procedure was used to measure the perceptual similarity structure of regional dialect variation in the United States. In two experiments, participants listened to a set of short English sentences produced by male talkers only (Experiment 1) and by male and female talkers (Experiment 2). The listeners were instructed to group the talkers by regional dialect into as many groups as they wanted with as many talkers in each group as they wished. Multidimensional scaling analyses of the data revealed three primary dimensions of perceptual similarity (linguistic markedness, geography, and gender). In addition, a comparison of the results obtained from the free classification task to previous results using the same stimulus materials in six-alternative forced-choice categorization tasks revealed that response biases in the six-alternative task were reduced or eliminated in the free classification task. Thus, the results obtained with the free classification task in the current study provided further evidence that the underlying structure of perceptual dialect category representations reflects important linguistic and sociolinguistic factors.

1. Introduction

Sociolinguistic variation is one source of phonetic variability in the speech signal and provides listeners with information about the talker's background, including his or her age, gender, ethnicity, and region of origin (Abercrombie, 1967). An understanding of the role of sociolinguistic variation in speech perception and production is essential for a complete theory of spoken language processing. The growing field of sociophonetics draws on experimental methods developed in acoustic phonetics and speech science to investigate the role of this source of information in the production, perception, and cognitive representation of linguistic variation (Thomas, 2002; Foulkes & Docherty, 2006).

Several recent studies on the perception of sociophonetic variation have used speech samples produced by talkers from different regions of the United States and the Netherlands to examine naïve listeners' identification and categorization of different linguistic varieties. In the United States, Clopper and Pisoni (2004b) asked naïve listeners to categorize unfamiliar male talkers by regional dialect using sentence-length materials in a six-alternative forced-choice categorization task. The talkers represented six regional varieties of American English. While overall accuracy was only 31%, categorization performance was reliably above chance. In addition, the large number of errors produced by the listeners provided sufficient data for an analysis of the perceptual similarity of the dialects. Using the stimulus–response confusion matrices from the six-alternative task, Clopper and Pisoni

(2004b) conducted a clustering analysis to explore the listeners' perceptual dialect categories. The results of our analysis revealed three main dialect categories: Northeast, South, and Midwest/West. Clopper, Conrey, and Pisoni (2005) replicated these findings with a set of female talkers and a mixed set of male and female talkers.

In the Netherlands, Van Bezooijen and Gooskens (1999) asked naïve adult listeners to categorize unfamiliar male talkers by country (the Netherlands or Belgium), region, and province of origin using samples of narrative speech. The Dutch listeners were able to categorize the Dutch talkers with 90%, 60%, and 40% accuracy at the country, region, and province levels, respectively. Van Bezooijen and Gooskens (1999) did not explicitly examine the error patterns produced by their listeners in the dialect categorization task; however, the high levels of accuracy in the country and region conditions suggest that many of the confusions were between two or more provinces within the same region. Van Bezooijen and Ytsma (1999) replicated these findings with a set of female Dutch speakers.

One of the primary limitations of these earlier studies is that the dialect regions and response labels were provided to the listeners by the experimenters. In order to reduce the constraints imposed by the verbal labels given to the listeners in a closed-set forced-choice task, the present study explored the perceptual structure of regional varieties of American English using an auditory free classification procedure (Imai, 1966; Imai & Garner, 1965). The use of a free classification task allowed the participants to create groups based on their own perceptual categories, without imposing an a priori geographic structure on their response categories. The free classification task has been used by perceptual dialectologists to explore naïve participants' beliefs about regional variation. For example, Tamasi (2003) gave participants a stack of index cards with the names of the 50 states and asked them to group the cards by how people talk in each state. Participants varied in their classification strategies, creating as few as five groups and as many as 35, with a mean of 14. These results suggest that naïve participants may have a large range of labels and regions that they consider to be culturally and linguistically salient. The current study was designed to explore this fine-grained knowledge in an explicit auditory dialect classification task.

1.1. Linguistic experience and the perception of dialect variation

One factor that has repeatedly been shown to affect the perception of dialect variation is the linguistic experience and developmental history of the participants. In one study, Clopper and Pisoni (2004a) recruited two groups of listeners to participate in a six-alternative forced-choice dialect categorization task. The listeners in one group, the “mobile” group, had lived in at least three different states at the time of testing. The second group of listeners, the “non-mobile” group, had all lived only in Indiana at the time of testing. A comparison of the categorization performance by these two groups revealed that the mobile listeners were more accurate than the non-mobile listeners in categorizing the unfamiliar talkers by regional dialect. In addition, a posthoc analysis comparing listeners who had lived in each dialect region to listeners who had not lived there revealed a significant difference in performance for residents and non-residents. Specifically, the residents of each region performed better than the non-residents in categorizing talkers from that dialect region, suggesting that direct exposure to different dialects as a result of residential history affects categorization accuracy. A clustering analysis of the stimulus–response confusion matrices for each listener group revealed the same three clusters reported by Clopper and Pisoni (2004b): Northeast, South, and Midwest/West. However, the mobile listeners tended to perceive greater differences between geographically contiguous regions, such as the North and North Midland or the North Midland and South Midland, than the non-mobile listeners did.

In another study, we (Clopper & Pisoni, 2006a) also found significant differences in perceptual dialect similarity due to the linguistic experience of the listeners. Four groups of listeners participated in a six-alternative forced-choice dialect categorization task. The listener groups represented two variables related to residential history: geographic mobility (mobile and non-mobile) and location (North and Midland). While all four groups performed the categorization task with the same degree of accuracy, a clustering analysis on the stimulus–response confusion matrices revealed significant effects of both geographic mobility and location. First, the mobile and non-mobile Northerners differed in their perception of the Northern talkers; the non-mobile Northerners perceived them as less distinctive from the other dialects than the mobile Northern listeners. That is, the listeners who moved to the Northern dialect region after having lived in another dialect region perceived the Northern dialect as more distinctive than the listeners who had only lived in the North. Second, the overall perceptual similarity structure of the dialects for the mobile North and the Midland listeners were identical, but the mobile North listeners perceived greater distinctiveness between the Southern and Mid-Atlantic talkers and the Midland and Western talkers than the Midland listeners did.

Finally, Williams, Garrett, and Coupland (1999) examined the perceptual categorization of dialect variation by adolescent boys in Wales. They asked young male listeners from six different regions of Wales to categorize regional varieties of Welsh English using spontaneous speech samples produced by a different set of adolescent boys from the same six cities. Overall categorization performance by the adolescents was 30% correct. However, they correctly categorized 45% of the talkers who came from their own region, but only 24% of the talkers who were from a different region. Williams et al. (1999) also collected categorization responses to the same stimuli from a set of Welsh schoolteachers, whose performance was much higher than the Welsh children's performance (52% correct for the adults vs. 30% for the children). The authors attributed the better performance by the adults to their greater familiarity with the varieties being tested as a result of more extensive travel experiences.

Region of origin and geographic mobility have also been found to affect the perception of dialect-specific variants in vowel perception tests. For example, Labov and Ash (1998) showed that listeners from Birmingham, Alabama, were able to more accurately identify Southern shifted vowels than listeners from either Chicago or Philadelphia. More recently, Evans and Iverson (2004a) found that listeners with limited exposure to different regional varieties of British English were less able to match target stimuli to the dialect of a preceding carrier phrase than listeners with greater exposure to variation due to geographic mobility. Rakerd and Plichta (2003) reported similar results for listeners from the United States with American English stimulus materials.

Taken together, all of these studies showed an effect of the location of the listeners, with a tendency for greater perceptual distinctiveness and categorization accuracy for local varieties. The one exception to this pattern was the Northern listeners, who were found to misperceive their own dialect (Clopper & Pisoni, 2006a). The studies by Williams et al. (1999), Evans and Iverson (2004a), and our earlier studies (Clopper & Pisoni, 2004a, 2006a) also revealed evidence of an effect of geographic mobility. The mobile listeners performed more accurately overall and perceived greater distinctiveness between the different dialects than the non-mobile listeners.

The present set of experiments was designed to further explore the role of residential history (specifically, geographic mobility and region of origin) on dialect perception using an unconstrained free classification task (Imai & Garner, 1965). The free classification task allowed us to examine naïve listeners' perceptual dialect categories and the perceptual

similarity of talkers with respect to regional dialect without providing a fixed number of categories or labels. We predicted that the free classification task would provide further insights into naïve listeners' perceptual representations of dialect variation by allowing them to create their own categories based on their perception of speech samples. By manipulating the residential history of the listeners we were also able to explore the role of linguistic experience and dialect familiarity in the perception and representation of dialect variation. The results of this study have implications for models of linguistic and sociolinguistic representations, which must account for both talker-specific aspects of spoken language processing, such as regional dialect, and the role of listener-specific experience, such as residential history, in the perception of linguistic and sociolinguistic information in speech.

Experiment 1 was a pilot study of the free classification paradigm using the same stimulus materials produced by male talkers as in our earlier forced-choice categorization experiment (Clopper & Pisoni, 2004b) to allow comparison of the results of the free classification task and the earlier categorization research on dialect perception. Like the listeners in the earlier study, the listeners in Experiment 1 were mixed with respect to their residential histories. Experiment 2 used the free classification method developed in Experiment 1 with the mixed male and female stimulus materials used by Clopper and Pisoni (2006a). In addition, four groups of listeners with different residential histories were recruited for participation in Experiment 2. The results of Experiment 2 were therefore directly comparable to the results obtained from the six-alternative forced-choice categorization task with four comparable sets of listeners (Clopper & Pisoni, 2006a).

Based on the findings obtained from our earlier forced-choice categorization studies (Clopper & Pisoni, 2004a, b, 2006a), we had several specific predictions regarding the outcome of the free classification experiments. First, given that naïve listeners appear to have three perceptual dialect categories, we predicted that they would create a relatively small number of groups of talkers in the free classification task. Second, we expected geographic mobility and location to affect performance, particularly with respect to the perception of a high degree of similarity between the Northern and Midland talkers for the non-mobile Northern listeners. Finally, we predicted that the mobile listeners would have developed more well-defined perceptual dialect categories as a result of their experience with multiple dialects and would therefore be able to more accurately distinguish talkers of different dialects and create more groups of talkers.

2. Experiment 1

2.1. Methods

2.1.1. Listeners—Twenty-five listeners were recruited from the Indiana University community for participation in this experiment. Prior to the analysis, data from three participants were removed: two reported a history of a hearing or speech disorder at the time of testing and one was substantially older than the other participants. The remaining 22 listeners were between the ages of 18 and 25 years old. The majority of the listeners ($N = 18$) were freshmen or sophomores at the time of testing. They were all monolingual native speakers of American English with no reported history of a hearing or speech disorder. Both parents of each participant were also native speakers of English. The residential history of the participants varied, but most ($N = 12$) were from the Midwest. Of the remaining listeners, 8 had lived in more than one dialect region, 1 was from the South, and 1 was from the West. The participants received \$8 for their service.

2.1.2. Talkers—Sixty-six male talkers from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986) produced the stimulus materials used in Experiment 1. The 66 talkers were between the ages of 20 and 29 at the

time of recording, with eleven talkers from each of six dialect regions in the United States: New England, North, North Midland, South Midland, South, and West.¹ The set of talkers used in the current experiment was identical to the set used in the previous forced-choice dialect categorization experiments by Clopper and Pisoni (2004a, b).

The six dialects represented by the talkers in Experiment 1 differed in terms of a number of segmental properties. Clopper and Pisoni (2004b) conducted an acoustic-phonetic analysis of the speech of the 66 talkers used in the current experiment and found a number of acoustic-phonetic variables that distinguished the six different dialects. The constraints of the TIMIT corpus limited our acoustic analysis to variables contained in the sentences, “She had your dark suit in greasy wash water all year.” and “Don’t ask me to carry an oily rag like that.” The characteristic properties of each dialect for this set of talkers are shown in Table 1. It should be noted, however, that additional acoustic-phonetic variables including other vowels and consonants, prosody, and voice quality might systematically vary across these dialects.

2.1.3. Stimulus materials—The stimulus materials consisted of one novel sentence per talker, for a total of 66 different sentences. The sentences were selected to contain dialect-specific vowel shifts, so that appropriate dialect variation was present in the stimulus materials. For example, the sentences for the Northern talkers typically contained a shifted /æ/ or /ɑ/, whereas the sentences for the Southern talkers typically contained a fronted /u/ or monophthongal /ay/.

The stimulus materials presented to the participants were 66 identical solid blue squares, each outlined with a medium gray box. Each square was linked to a different sound file containing one of the novel sentences. The original TIMIT wav files were edited to include only the speech material and the mean amplitude of each sentence was leveled to 55 dB using Level16 (Tice & Carrell, 1998).

2.1.4. Procedure—The participants were seated at personal computers equipped with a two-button computer mouse and a set of Beyerdynamic DT100 headphones. On the computer screen, the participants saw 66 blue squares arranged in columns on the left and a 20×20-cell grid on the right. The participants could listen to each sound file by clicking on the blue square with the right mouse button. The sound file was presented at approximately 70 dB SPL over the headphones. The participants could move the blue square around the screen by clicking and dragging it with the left mouse button.

The participants were told that each of the blue squares on the left side of the screen represented a different talker and that the talkers came from different parts of the United States. They were asked to group the talkers based on where they thought the talkers were from. The listeners were allowed to make as many groups as they wanted with as many talkers in each group as they wished. They did not have to put the same number of talkers in each group and they could listen and move the talkers around as many times as they wanted. No time limit was imposed on the procedure and the participants were simply instructed to tell the experimenter when they were finished.

¹The dialect region labels used in Experiment 1 refer to the labels provided with the original TIMIT corpus. While these dialect regions do not correspond to the dialect regions in Experiment 2 or to current sociolinguistic descriptions of regional variation in the United States, additional information about the residential history of the TIMIT talkers is no longer available, so reassignment of the talkers to more appropriate dialect groups was not possible.

2.2. Results

On average, the participants made 10 groups of talkers, with a range of 3–30 and a median of 7. The mean number of talkers per group was 9.36, with a range of 1–34 and a median of 4.

2.2.1. Perceptual dialect similarity—The perceptual similarity structure of the six dialect regions was extracted from the free classification data using an additive clustering analysis which produced graphical models of perceptual similarity in tree form (Cortier, 1982; Sattath & Tversky, 1977). First, a 6×6 dialect similarity matrix was constructed from the free classification data. The matrix reflected the similarity of the dialects as measured by pairwise comparisons of the talkers within each group created by each listener. The value of each cell in the diagonal of the matrix was equal to the number of times that two talkers from the same dialect were put in the same group, summed across all of the listeners. The value of each cell in the off-diagonals of the matrix was equal to the number of times one talker from one dialect and one talker from another dialect were put in the same group, summed across all listeners.

To obtain a graphical representation of the perceptual similarity of the six dialects, the 6×6 dialect similarity matrix was submitted to ADDTREE (Cortier, 1982), an additive clustering analysis. The clustering solution produced by ADDTREE is shown in Fig. 1. In this representation, perceptual similarity is indexed by the sum of the least number of vertical branches connecting any two-dialect nodes and horizontal distance is irrelevant. The clustering analysis revealed three main perceptual clusters: New England, South, and Midwest/West.

2.2.2. Perceptual talker similarity—A multidimensional scaling analysis was then carried out to explore the perceptual similarity of the talkers in Experiment 1. A 66×66 talker matrix was constructed from the free classification data by assigning to each cell the total number of times a given pair of talkers was put in the same group across all of the listeners. Thus, the 66×66 talker matrix reflected the pairwise similarity of all of the talkers. The resulting talker matrix was submitted to a multidimensional scaling analysis and the two-dimensional solution was selected for interpretation and discussion. The dimensions in the two-dimensional space were highly interpretable and the reduction in stress between the two- and three-dimensional solutions was relatively small.

Fig. 2 shows the results of the multidimensional scaling analysis. In this figure, each symbol represents one of the 66 talkers. The three categories that were revealed by the clustering analysis in Fig. 1 are also evident in this representation of perceptual talker similarity. The Southern and South Midland talkers are almost all located in the upper left-hand portion of the space, the New England talkers are almost all in the lower right-hand quadrant, and the rest of the talkers are in the upper right-hand region.

Rotation of the space approximately 60° allows for the interpretation of the two perceptual dimensions. From the upper right to the lower left is a perceptual dimension related to linguistic markedness, with the marked dialects on the bottom left and the unmarked dialects on the upper right. The term “linguistic markedness” refers here to the degree to which a given dialect contains phonological variants that are different from the other dialects (Milroy, 2002). These differences are assessed in both historical terms, so that the result of recent changes are assumed to be marked, and synchronic comparative terms, so that variants that are exhibited by only a single dialect are assumed to be marked. For example, /u/ fronting in the South Midland is marked because it is a relatively recent innovation (Thomas, 2001), whereas r-lessness in New England is marked because it is a relatively rare property of American English. In addition, the relative markedness of a given dialect is

assumed to be based on objective descriptions of its phonological system and not the relative salience of specific phonological variants for naïve listeners (Labov, 2001). For American English, objective descriptions of regional variation tend to focus on the vowel system, which means that the interpretation of relative markedness is also based almost entirely on vowel-based differences. In the current experiment, the Southern and New England dialects are linguistically marked because they have more features that distinguish them from the other regional varieties. The North Midland and Western dialects, on the other hand, are linguistically unmarked because they have fewer features that distinguish them from the other dialects of American English (see Table 1). The second dimension in Fig. 2 is orthogonal to the first and distinguishes between Southern varieties at the upper left and Northern varieties at the lower right. Thus, the two perceptual dimensions that are relevant to the perceptual similarity of the talkers in Experiment 1 are linguistic markedness and geography.

2.3. Discussion

The participants in this experiment exhibited a range of free classification strategies, with some listeners making as few as three groups and others as many as 30. The mean number of groups produced was 10, which suggests that listeners can make a relatively large number of fine distinctions based on regional dialect using a free classification procedure. When the listeners were asked to provide labels for the groups of talkers they had made after completing the task, they typically provided a list of geographic regions, suggesting that they had followed the instructions and created groups based on their perception of regional varieties of American English. Therefore, while the results of the earlier six-alternative forced-choice categorization tasks seemed to suggest that naïve listeners have only three broad dialect categories, the present results are more consistent with the findings from Tamasi's (2003) free classification study and suggest that listeners can make finer distinctions between regional varieties when specific labels are not imposed on the task a priori by the experimenter.

Despite the fine-grained classifications made by individual participants, however, the clustering analysis of the aggregate data again revealed three broad dialect categories: New England, South, and Midwest/West. These three broad perceptual categories correspond directly to the perceptual categories revealed by clustering analyses of the confusion matrices in the earlier six-alternative forced-choice categorization tasks using the same set of stimuli from the TIMIT corpus (Clopper & Pisoni, 2004a, b). This result provides converging evidence for the validity of the previous results and confirms that a free classification task can be used to measure the perceptual similarity of regional dialects using spoken sentences.

The multidimensional scaling analysis also revealed several novel findings. When the perceptual distances between the 66 talkers in the current study were plotted in a two-dimensional space, the dimensions, which emerged, corresponded to linguistic markedness and geographic location. In particular, the markedness dimension distinguished marked dialect regions, like New England and the South, from unmarked regions, like the North Midland and the West. The geographic dimension, on the other hand, distinguished the Northern varieties, like New England and the North, from the Southern varieties, like the South and South Midland.

The success of the free classification paradigm in eliciting interpretable results is particularly compelling given the nature of the stimulus materials used in this study. Recall that each talker produced a different novel sentence, so that the listeners were required to make their judgments about dialect similarity in the absence of identical linguistic content. Thus, the task did not allow the listeners to focus on a small set of fixed linguistic features in

making their classifications, but instead required them to use abstract representations of the talkers and the dialect groups based on short, non-identical samples of speech.

The results obtained from this initial free classification experiment were promising because they were consistent with previous research on the perception of dialect variation, particularly with respect to the underlying perceptual similarity structure of regional dialects in the United States. In addition, the multidimensional scaling analysis of the classification data provided new insights into the relevant perceptual dimensions of variation for naïve listeners. The second experiment was designed to explore the effects of the residential history of the listeners, particularly with respect to geographic mobility and location, on free classification performance. In addition, Experiment 2 used stimulus materials from the Nationwide Speech Project (NSP) corpus (Clopper & Pisoni, 2006b), which more accurately reflects current regional variation in the United States. Thus, Experiment 2 was also designed as a replication of Experiment 1 using a different set of stimulus materials.

3. Experiment 2

3.1. Methods

3.1.1. Listeners—One hundred and six listeners were recruited from the Indiana University community for participation in this experiment. Prior to analyzing the data, 19 participants were removed for the following reasons: seven could identify one or more of the talkers by name, one had a parent who was a non-native speaker of English, and 11 did not perform the task as instructed.² The remaining 87 listeners were all monolingual native speakers of American English with no reported history of a hearing or speech disorder. Both parents of each listener were also native speakers of English. The listeners ranged in age from 18 to 25 years old, but most ($N = 57$) were freshmen or sophomores at the time of testing. The listeners received \$8 for their participation.

The listeners were assigned to four different groups based on their residential history. Twenty-one listeners had lived only in the Midland dialect region and they formed the non-mobile Midland group. Twenty-two listeners had lived only in the Northern dialect region prior to attending college at Indiana University in Bloomington and they comprised the non-mobile North group. Forty-four listeners had lived in at least two different dialect regions before the age of 18. Twenty-two of these mobile listeners had parents living in the Midland dialect region at the time of testing and they comprised the mobile Midland group. The remaining twenty-two mobile listeners had parents living in the Northern dialect region at the time of testing and they comprised the mobile North group.

3.1.2. Talkers—Forty-eight talkers were selected from the NSP corpus (Clopper & Pisoni, 2006b) for use in Experiment 2. The talkers included four males and four females from each of six dialect regions in the United States: New England, Mid-Atlantic, North, Midland, South, and West. This set of talkers is the same as those used in the forced-choice categorization experiment described by Clopper and Pisoni (2006a).

The six regional dialects included in the NSP corpus were selected because talkers from those regions exhibit systematic differences in their vowel productions. The vowel system of the Northern dialect is characterized by the Northern Cities Chain Shift, which involves a clockwise rotation of the low and low-mid vowels, beginning with the raising and fronting

²Data were excluded for those participants whose responses revealed a greater proportion of between-dialect similarity than proportion of within-dialect similarity. This is equivalent to the proportion of errors being greater than the proportion of correct responses, given all possible responses. It is impossible to determine if these participants were performing the task to the best of their abilities and were simply unable to accurately group the talkers by dialect or if their performance was due to lack of attention to the task. This strict exclusion criteria was used to ensure that the data reflected the best efforts of the participants.

of /æ/ (Labov, 1998). The New England dialect also exhibits some properties of the Northern Cities Chain Shift, including /æ/ raising and fronting, particularly in Western New England (Boberg, 2001). The vowel system of the Southern dialect includes shifts in both the back vowels and the front vowels. The high and mid back vowels are fronted, while the front high and mid tense vowels are centralized and the front high and mid lax vowels are peripheralized (Labov, 1998). Southern speech also exhibits the monophthongization of /ay/ and /oy/ (Thomas, 2001). The vowel system in the “Third Dialect” of American English, which includes New England, the Midland, and the West, shows a merger of the low-back vowels /ɔ/ and /ɑ/ (Labov, 1998). Midland speech also exhibits /u/ and /ow/ fronting (Labov, Ash, & Boberg, 2005) and Western speech also exhibits /u/ fronting (Labov et al., 2005; Thomas, 2001). Finally, the Mid-Atlantic dialect is characterized by the raising of /ɔ/ and the raising of /æ/ in certain lexical contexts (Labov, 1994; Thomas, 2001). The vowel productions of the talkers included in the NSP corpus are discussed in more detail by Clopper, Pisoni, and de Jong (2005). It should also be noted that unlike the New England talkers in Experiment 1, who were all r-less, the New England talkers from the NSP corpus used in Experiment 2 were all r-ful.

3.1.3. Stimulus materials—For each talker, one novel sentence was selected, for a total of 48 novel sentences. The sentences were taken from the Speech Perception in Noise (SPIN) test (Kalikow, Stevens, & Elliott, 1977). The stimuli were presented using the same methods as in Experiment 1, except that the sound files were leveled to 67 dB using Level16 (Tice & Carrell, 1998).

3.1.4. Procedure—The procedure was identical to that used in Experiment 1, except that only 48 stimulus items were presented to the listeners. Given that both male and female talkers were included in Experiment 2, an additional instruction was also given to the listeners. They were told that they could put males and females in the same group if they thought that they were from the same part of the country.

3.2. Results

A summary of the response patterns in the free classification task for each of the four listener groups is shown in Tables 2 and 3. Table 2 displays descriptive statistics on the number of talker groups created by each of the listener groups. Table 3 shows descriptive statistics on the number of talkers per group for each of the listener groups.

Overall, the listeners produced an average of 8.48 groups, with a range of 3–23 and a median of 8. A one-way ANOVA on the number of talker groups created by each listener group (mobile North, mobile Midland, non-mobile North, non-mobile Midland) was not significant. An inspection of Table 2 suggests, however, that mobility may be an important variable with respect to the observed response patterns. A *t*-test confirmed that the mobile listeners created significantly more groups ($M = 9.5$) than the non-mobile listeners ($M = 7.5$; $t(85) = 2.24, p < .05$). *t*-tests comparing the mobile and non-mobile Northerners and the mobile and non-mobile Midland listeners were not significant. In addition, location was not significant by a *t*-test.

With respect to the number of talkers per group, the listeners created groups containing an average of 7.08 talkers, with a range of 1–38 and a median of 4. A one-way ANOVA on the number of talkers per group for each listener group (mobile North, mobile Midland, non-mobile North, non-mobile Midland) was not significant. *t*-tests comparing the mobile and non-mobile listeners and the Midland and Northern listeners were also not significant, suggesting that geographic mobility and location were not significant variables in determining the number of talkers per group produced by the listeners. Thus, while the

geographic mobility of the listeners had an affect on their free classification strategy, location did not. In addition, the effect of mobility was only revealed by the comparison of the number of groups of talkers across all four-listener groups.

The listeners' performance was also assessed in terms of their ability to accurately group the talkers by dialect in the free classification task. First, for each listener, a “percent correct” score was calculated as the number of times talkers from the same dialect were grouped together out of the total number of possible same-dialect pairings. Second, a “percent error” score was calculated for each listener reflecting the number of times talkers from different dialects were grouped together out of the total number of possible different-dialect pairings. Table 4 shows the mean percent correct and error scores for each of the four listener groups.

One-way ANOVAs on the percent correct and percent error scores were computed with listener group (mobile North, mobile Midland, non-mobile North, non-mobile Midland) as the factor. The results of both ANOVAs were not significant, suggesting no differences between the groups in terms of their overall accuracy.

3.2.1. Perceptual dialect similarity—Perceptual dialect similarity was assessed using the clustering techniques described in Experiment 1. To compare the perceptual similarity spaces of the four listener groups, a 6×6 dialect similarity matrix was constructed for each listener group. The four matrices were then submitted to ADDTREE, the additive clustering model (Corter, 1982). The resulting graphical models of similarity are shown in Fig. 3. As in Fig. 1 above, perceptual dissimilarity is represented as a function of the lengths of the least number of vertical bars connecting any two dialect nodes and horizontal distance is irrelevant. The overall structure of the similarity spaces of the dialects is fairly consistent across the four listener groups. In particular, for all four listener groups, the Mid-Atlantic and the South are the most distinctive dialects, and the Midland and Western dialects are perceived as highly similar.

However, some differences in the similarity spaces between the listener groups are clearly visible in the clustering solutions in Fig. 3. The perceptual similarity spaces for the mobile listeners more accurately reflect the phonological characteristics of the dialects, with a tight clustering of the New England, Midland, and Western dialects and greater perceptual distances between the Northern, Southern, and Mid-Atlantic varieties. The primary difference in structure between the mobile North and mobile Midland listeners is that the mobile Northern listeners perceived the Northern talkers as more similar to Labov's (1998) “Third Dialect” talkers than the mobile Midland listeners did.

The perceptual dialect similarity structures for the non-mobile listeners are less closely related to the phonological properties of the different dialects. For the non-mobile Northern listeners, the most notable feature in the clustering solution is the close relationship between the Northern talkers and the New England, Midland, and Western talkers. This relationship is similar to that reported by Clopper and Pisoni (2006a) for the six-alternative forced-choice task and may reflect the non-mobile Northern listeners' inattention to the Northern Cities Chain Shift. The non-mobile Midland listeners produced the most unexpected similarity structure, particularly with respect to the high degree of perceptual similarity between the Southern talkers and the Midland and Western talkers.

3.2.2. Perceptual talker similarity—To assess the perceptual similarity of the 48 talkers in Experiment 2, a multidimensional scaling analysis was conducted. As in Experiment 1, a talker similarity matrix was constructed by summing over all of the listeners so that each cell of the matrix indicated the total number of times that that pair of talkers was grouped

together in the free classification task. The 48×48 talker matrix was then submitted to a multidimensional scaling analysis.

A three-dimensional solution was selected for interpretation and discussion, based on the relatively high interpretability of the dimensions and the fact that stress was much greater for the two-dimensional solution and not greatly reduced for the four-dimensional solution. Figs. 4 and 5 show the similarity space produced by the multidimensional scaling analysis in three dimensions. Dimensions 1 and 2 are plotted against each other in Fig. 4 and Dimensions 1 and 3 are plotted against each other in Fig. 5. The filled symbols represent male talkers and the open symbols represent female talkers. For the West, plain X's indicate females and boxed X's indicate males.

Unlike the multidimensional scaling analysis obtained in Experiment 1, the three dimensions that emerged from the analysis in Experiment 2 are interpretable without rotation. Dimension 1 corresponds to markedness, with the linguistically marked dialects on the right and the linguistically unmarked dialects on the left. Most of the Mid-Atlantic and Southern talkers fall on the positive side of Dimension 1 and most of the Midland, New England, and Western talkers fall on the negative, or unmarked, side. Dimension 2 is related to gender, with nearly all of the male talkers above zero and most of the female talkers below zero. Finally, Dimension 3 can be interpreted geographically, with the South at the top and the North at the bottom. All of the Southern talkers have positive values on Dimension 3, whereas all of the Mid-Atlantic and most of the New England and Northern talkers have negative values.

The perceptual similarity space described above and shown in Figs. 4 and 5 is the result of combining the responses of all of the listeners in all four of the listener groups. To assess differences in perceptual similarity due to residential history, one talker similarity matrix was constructed for each listener group, for a total of four 48×48 talker matrices. These four matrices were then submitted to an Individual Differences Scaling (INDSCAL) multidimensional scaling analysis (Carroll & Chang, 1970). The INDSCAL model returns a single group space and subject weights for each matrix. The group space in this case models the perceptual similarity of the 48 talkers and was fixed to be identical to the space produced by the multidimensional scaling analysis based on all of the listeners' data and plotted in Figs. 4 and 5.

The subject weights indicate the relevance of each of the three dimensions for each listener group. The subject weights produced by the INDSCAL analysis are proportional to the overall model fit for each subject. To compare the weights across the listener groups, overall model fit has been factored out by normalizing the subject weights for each group so that they sum to 1. Table 5 shows the normalized weights for each listener group for each dimension.

Overall, the markedness dimension (Dimension 1) received the highest weights, followed by gender (Dimension 2) and geography (Dimension 3). Across all four listener groups, markedness was the most relevant factor in assessing the similarity of the talkers in the free classification task, followed by gender and then geography. The Northern listener groups showed slightly more attention than average to the markedness dimension, while the Midland listener groups showed slightly more attention than average to the gender dimension. The non-mobile Northern listeners were most attentive to the geographic dimension, relative to the other listener groups. The listener group differences were quite small, however, suggesting a similar set of relevant perceptual dimensions and attentional weightings across all four groups.

3.3. Discussion

Overall performance measured in terms of categorization accuracy was slightly higher in the free classification task (28%) than in the earlier forced-choice task (26%; Clopper & Pisoni, 2006a), but the error scores calculated from the free classification data revealed a large percentage of errors (19%) as well. Classification performance measured in terms of accuracy and error scores did not differ across the four listener groups. These results are consistent with the results of the six-alternative forced-choice categorization task obtained by Clopper and Pisoni (2006a), who also did not observe any main effects of residential history on categorization accuracy. However, in the free classification task, residential history did affect the classification strategies of the listeners. In particular, the mobile listeners created more talker groups on average than the non-mobile listeners. This finding suggests that personal experience with different regional varieties led to more detailed perceptual categories for the mobile listeners.

The perceptual similarity of the six regional dialects was revealed by the clustering analysis. While the overall similarity spaces were consistent across all four listener groups, with Northeast, South, and Midwest/West categories, both the geographic mobility and location of the listeners affected their performance in the free classification task. In particular, the perceptual similarity structure of the six dialects for the mobile listener groups was similar to what would be predicted based on an analysis of the phonological properties of the dialects; Mid-Atlantic and Southern are the most linguistically marked dialects and they were also the most perceptually distinctive. The Midland, Western, and New England dialects are the least marked and they clustered tightly together. The Northern dialect is also fairly marked phonologically, but the perception of the Northern talkers was affected by the listeners' location. The mobile Northern listeners tended to perceive the Northern talkers as more similar to the Midland, Western, and New England talkers, whereas the mobile Midland listeners heard a greater similarity between the Northern and Mid-Atlantic talkers.

The bias for Northern listeners to hear the Northern talkers as less marked is also evident in the perceptual similarity structure of the six dialects for the non-mobile Northern listeners. These listeners perceived the Northern talkers as highly similar to the New England talkers, as well as to the Midland and Western talkers. The poorer perception of the distinctive features of the Northern Cities Chain Shift by Northern listeners was also found in the clustering analysis based on the error patterns in the six-alternative forced-choice task, particularly for the non-mobile Northern listeners (Clopper & Pisoni, 2006a). Similar results have also been reported in explicit categorization studies examining the perceptual identification of shifted and unshifted vowels by Northern listeners (Niedzielski, 1999; Rakerd & Plichta, 2003).

Finally, the non-mobile Midland listeners showed an unexpected pattern of perceptual similarity with a high degree of similarity between the Southern and Midland talkers. This result is unusual in comparison to the other listener groups, who all perceived the Southern talkers as fairly distinct from the other dialects. This finding was also unexpected, based on the previous six-alternative forced-choice experiments (Clopper & Pisoni, 2004a, 2006a) which revealed both high levels of categorization accuracy for Southern talkers and perceptual distinctiveness of Southern talkers for non-mobile Midland listeners.

Thus, with respect to judgments of perceptual dialect similarity, both geographic mobility and location appear to be important factors in shaping naïve listeners' perceptions. In particular, geographic location is important for distinguishing between local dialects, particularly with respect to the Northern listeners' misperception of the Northern Cities Chain Shift. Mobility increases listeners' familiarity with other dialects, which leads to more perceptual categories and better discrimination of local dialects. Specifically, the mobile

Northern listeners perceived a greater difference between the North and the Midland than the non-mobile Northerners did. Similarly, the mobile Midland listeners perceived a greater difference between the Midland and the South than the non-mobile Midland listeners did.

The perceptual similarity structures revealed by the free classification task differ in several important ways from those that emerged from the analyses of the six-alternative forced-choice confusion data using the NSP stimulus materials. First, while Clopper and Pisoni (2006a) found that New England was clustered most closely with the Mid-Atlantic region in the clustering solutions for all of the listener groups in the six-alternative forced-choice task, the free classification data from the present study show a high degree of perceptual similarity between New England and the other “Third Dialect” regions, including the Midland and the West. This difference in the perceptual dialect similarity structures across the two tasks can be accounted for by further examination of the forced-choice categorization data. In particular, the perceptual similarity of the New England and Mid-Atlantic talkers was asymmetric in the forced-choice categorization task; Mid-Atlantic talkers were categorized as New England more often than New England talkers were categorized as Mid-Atlantic. Based on these findings, we suggested that the listeners had only one category for Northeastern speech and that this category reflected the phonological variables associated with the Mid-Atlantic talkers. This interpretation is consistent with the pattern of perceptual similarity obtained with the current free classification task.

A second major difference between the perceptual similarity spaces produced from the two different tasks is the perceived similarity between the Midland and Southern talkers for the Midland listeners. In the six-alternative task, we found that the Southern talkers were categorized more accurately than most of the other talker groups by all four listener groups (Clopper & Pisoni, 2006a). In addition, the Southern dialect was found to be distinctive from the other dialects in the clustering analyses based on the stimulus–response confusions for all of the listener groups. In the free classification experiment, however, the non-mobile Midland listeners behaved like the non-mobile Northern listeners because they failed to distinguish talkers from their own region from talkers from a neighboring region. Specifically, while the non-mobile Northern listeners perceived a greater similarity between the Northern and Midland talkers than the other listener groups, the non-mobile Midland listeners perceived a greater similarity between the Midland and Southern talkers than the other listener groups. This finding suggests that the response labels provided in the forced-choice categorization experiment (Clopper & Pisoni, 2006a) forced the non-mobile Midland listeners to make distinctions between Midland and Southern talkers that they did not find relevant in the free classification task.

The perceptual similarity structure obtained in this experiment also differed from the structure obtained in Experiment 1 using the TIMIT stimulus materials, particularly with respect to the perception of the New England talkers. Recall that the New England talkers in Experiment 1 were all *r*-less, which has been shown to be a very salient perceptual cue for naïve listeners (Clopper & Pisoni, 2004b). However, none of the New England talkers in Experiment 2 were *r*-less, which means that the listeners had to rely on other phonological properties of their speech, such as vowel quality, to distinguish the New England talkers from the other talkers in this task. This difference in *r*-lessness between the two groups of talkers is probably the major factor responsible for the relatively greater perceptual distinctiveness of the New England talkers in Experiment 1 than Experiment 2.

The multidimensional scaling analysis used in the current experiment revealed a perceptual talker similarity space with three dimensions: linguistic markedness, gender, and geography. Together, the markedness and geography dimensions define a perceptual space in which the three broad dialect categories can be distinguished (see Fig. 5). The Southern dialect

occupies the marked Southern quadrant of the space, the Mid-Atlantic and Northern dialects occupy the marked Northern quadrant of the space, and the Midland, Western, and New England dialects occupy the unmarked Northern quadrant. The unmarked Southern quadrant is fairly empty, because Southern varieties of American English are typically marked.

The fact that gender emerged as a relevant perceptual dimension in the analysis is interesting because the listeners were told to ignore gender in making their groups. This instruction may be difficult to follow, however, because gender and dialect are known to interact. For example, women typically lead phonological change, which means that women might produce more of some variants than men (Labov, 1990). On the other hand, women also tend to avoid stigmatized forms, which means that they might produce fewer of some other variants than men (Labov, 1990). Thus, the fact that gender emerges as an important perceptual dimension may reflect the listeners' sensitivity to the complex interaction between gender and regional dialect in speech perception. Alternatively, it may be that listeners cannot ignore salient talker differences, such as those related to gender, in perception tasks, even when they are explicitly instructed to attend to a different dimension of variability (see e.g., Clopper, Levi, & Pisoni, 2006; Mullennix & Pisoni, 1990).

Our finding that naïve listeners display sensitivity to variation in the linguistic (e.g., phonological) and indexical (e.g., gender or dialect) information in the speech signal is also consistent with previous findings which suggest that speech perception is a talker-contingent process (Pisoni, 1997). For example, Nygaard, Sommers, and Pisoni (1994) found that speech intelligibility in noise was more accurate when the talkers were familiar to the listeners than when they were unfamiliar, suggesting an integration of phonological and talker-specific information in speech perception. The results of the present study suggest that naïve listeners can integrate several different kinds of talker-specific and linguistic information in order to make perceptual judgments about the dialect of a given talker.

4. General discussion

Free classification procedures revealed several new insights about the underlying perceptual similarity structure of regional varieties of American English for naïve listeners. First, the listeners in both experiments made more groups of talkers than predicted, suggesting that they are able to make fine-grained distinctions between regional dialects of American English. Second, the clustering analyses revealed parallel patterns of similarity structure to those revealed by the earlier forced-choice tasks (Clopper & Pisoni, 2004b, 2006a). These findings provide converging evidence for the three primary perceptual dialects of American English: Northeast, South, and Midwest/West. Third, in Experiment 2, we found that New England talkers were perceived as being more similar to Midland and Western talkers, with whom they share several phonological properties, than in the previous forced-choice task (Clopper & Pisoni, 2006a), in which they were perceived as being most similar to their geographic neighbors from the Mid-Atlantic region. This result suggests that the labels provided by the experimenter in a forced-choice task may lead to some response biases that can be reduced by using a free classification task. Fourth, residential history was found to be an important factor in the performance of the listeners in Experiment 2. The mobile listeners perceived larger differences between geographically local dialects than the non-mobile listeners, suggesting that direct exposure to different regional varieties affects the perception and representation of dialect variation. Finally, the multidimensional scaling analyses revealed three primary underlying dimensions of perceptual organization of dialect variation: linguistic markedness, geography, and gender. All three of these perceptual dimensions have both linguistic and social significance, suggesting that naïve listeners build perceptual categories for regional variation based on appropriate sociolinguistic dimensions that reflect the phonological variation they have been exposed to.

Several questions remain, however, regarding the nature of cognitive dialect representations and the role of experience in developing these categories. First, we have defined dialect markedness with respect to phonological descriptions of variation in the sociolinguistic literature. However, descriptions of regional variation in the United States have focused almost exclusively on vowels and we know virtually nothing about variation in other phonological domains, such as consonants, prosody, or voice quality (Docherty & Foulkes, 1999). In addition, while Labov (2001) described three levels of perceptual salience for sociolinguistic variables, we do not have empirical data to show which variables are salient for naïve listeners and which are not. The results of the free classification experiments suggest that some variables, such as r-lessness in New England, are very salient properties for Midwestern listeners, whereas other variables, such as the Northern Cities Chain Shift, are not. Additional research is needed to explore the linguistic features that are central to judgments of dialect category membership and prototypicality as well as the range of linguistic features that are perceptually marked for naïve listeners.

Second, the dialects selected for inclusion in the present study were highly constrained in that we only included white regional varieties of American English. If the set of dialects were expanded to include ethnic varieties (e.g., African American Vernacular English), non-North American varieties (e.g., Scottish English), or foreign accented English, the results might provide additional insights into the perceptual structure of dialect representations. Given that the listeners in the present free classification study tended to perceive geographically local dialects as more similar and geographically distant dialects as more distinct, we would predict that the addition of a very different dialect might lead to an “us versus them” classification strategy in which the white North American varieties would be perceived as more similar and the ethnically or globally different dialects as more distinct. Alternatively, if we limited the set of dialects to local Midwestern varieties, we might find that the listeners make finer discriminations between local varieties. Additional research is needed to determine the role of the stimulus set in dialect classification behavior and the limits of perceptual dialect discrimination for geographically local varieties.

Third, while we made a distinction in Experiment 2 between mobile and non-mobile listeners, all of the listeners in our experiments had some exposure to regional dialect variation because they were all students at Indiana University at the time of testing. Evans and Iverson (2004b) reported significant changes in vowel production after 2 years in a multi-dialect university setting in Great Britain, but no significant differences in vowel perception over the same time period. However, the results of Experiment 2 suggest that geographic mobility prior to attending college affects the explicit classification of dialects and we would predict that the effects of mobility would be more pronounced for listener populations with more extreme residential histories (e.g., children of military personnel who move around frequently or non-mobile young adults who attend a local community college). Additional research is needed to examine the interactions between age, geographic mobility, and level of education in the representation of perceptual dialect similarity structure.

Taken together, these three remaining questions suggest an approach that adopts an exemplar-based representation of dialect variation (Clopper, 2004; Johnson, 1997). In particular, salient phonological properties are stored as exemplars of a given dialect, the salience of a given acoustic-phonetic cue in perception is partially determined by other available dialect information (i.e., by the available stimulus set), and the listener's experience with variation has a significant impact on the development of dialect categories. For example, geographic mobility provides naïve listeners with the opportunity to interact with talkers from different parts of the country and to accurately assign dialect labels to those individuals. Non-mobile listeners, however, have many fewer opportunities to make connections between talkers and their residential history and, by extension, their appropriate

dialect category label. That is, the mobile listeners were more likely to have accurately labeled exemplars of different varieties in long-term memory than the non-mobile listeners. This difference in direct exposure to talkers and their dialect labels therefore led to more accurate perceptual similarity between dialect categories for the mobile listeners than the non-mobile listeners (see Barsalou, 1985). An exemplar-based approach may provide a useful model of the representation of dialect variation and the interaction between linguistic and sociolinguistic information in speech perception.

The free classification task used in the present study is a promising new experimental methodology in the study of the perception of dialect variation by naïve listeners. The methodology could also be used for other perceptual studies in which the participants are asked to group a set of stimulus items in some way. Variations of the task allow the experimenter to impose structure on the number of categories or the spatial arrangement of the categories to further explore specific aspects of classification behavior. Similarly, the statistical methods applied to the free classification data (clustering and multidimensional scaling) can be applied to any dataset for which inter-stimulus distances can be quantified. For example, production data can be used to model dialect similarity (Nagy, Zhang, Nagy, & Schneider, 2005; Nerbonne & Heeringa, 2001) or language similarity. These methods therefore have applications beyond sociophonetics and may also be useful in studies of second language acquisition, historical linguistics, and linguistic typology.

Acknowledgments

This work was supported by NIH NIDCD T32 Training Grant DC00012 and NIH NIDCD R01 Research Grant DC00111 to Indiana University. The authors would like to thank Luis Hernandez and David Parsons for their assistance in the development of the experimental protocol and Robert Nosofsky for his advice on the statistical analyses.

References

- Abercrombie, D. Elements of general phonetics. Edinburgh University Press; Edinburgh: 1967.
- Barsalou LW. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1985; 11:629–654.
- Boberg C. The phonological status of Western New England. *American Speech*. 2001; 76:3–29.
- Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an *n*-way generalization of “Eckart-Young” decomposition. *Psychometrika*. 1970; 35:238–319.
- Clopper, CG. Linguistic experience and the perceptual classification of dialect variation. Indiana University; Bloomington: 2004. Doctoral dissertation
- Clopper CG, Pisoni DB. Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*. 2004a; 16:31–48.
- Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*. 2004b; 32:111–140.
- Clopper CG, Pisoni DB. Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change*. 2006a; 18:193–221.
- Clopper CG, Pisoni DB. The Nationwide Speech Project: A new corpus of American English dialects. *Speech Communication*. 2006b; 48:633–644.
- Clopper CG, Pisoni DB, de Jong K. Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*. 2005; 118:1661–1676. [PubMed: 16240825]
- Clopper CG, Conrey BL, Pisoni DB. Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology*. 2005; 24:182–206.
- Clopper CG, Levi SV, Pisoni DB. Perceptual similarity of regional dialects of American English. *Journal of the Acoustical Society of America*. 2006; 119:566–574. [PubMed: 16454310]

- Cortier JE. ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. *Behavior Research Methods and Instrumentation*. 1982; 14:353–354.
- Docherty, GJ.; Foulkes, P. Derby and Newcastle: Instrumental phonetics and variationist studies. In: Foulkes, P.; Docherty, GJ., editors. *Urban voices*. Edward Arnold; London: 1999. p. 47-71.
- Evans BG, Iverson P. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*. 2004a; 115:352–361. [PubMed: 14759027]
- Evans BG, Iverson P. Plasticity in speech perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*. 2004b; 117:2426.
- Fisher, WM.; Doddington, GR.; Goudie-Marshall, KM. The DARPA speech recognition research database: Specification and status; Proceedings of the DARPA speech recognition workshop; 1986. p. 93-99.
- Foulkes P, Docherty GJ. The social life of phonetics and phonology. *Journal of Phonetics*. 2006; 34:4.
- Imai S. Classification of sets of stimuli with different stimulus characteristics and numerical properties. *Perception and Psychophysics*. 1966; 1:48–54.
- Imai S, Garner WR. Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*. 1965; 69:596–608. [PubMed: 14304312]
- Johnson, K. Speech perception without speaker normalization: An exemplar model. In: Johnson, K.; Mullennix, JW., editors. *Talker variability in speech processing*. Academic Press; San Diego: 1997. p. 145-165.
- Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*. 1977; 61:1337–1351. [PubMed: 881487]
- Labov W. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*. 1990; 2:205–254.
- Labov, W. *Principles of linguistic change: Internal factors*. Blackwell; Malden, MA: 1994.
- Labov, W. The three dialects of English. In: Linn, MD., editor. *Handbook of dialects and language variation*. Academic Press; San Diego: 1998. p. 39-81.
- Labov, W. *Principles of linguistic change: Social factors*. Blackwell; Malden, MA: 2001.
- Labov, W.; Ash, S. Understanding Birmingham. In: Bernstein, C.; Nunnally, T.; Sabino, R., editors. *Language variety in the south revisited*. Alabama University Press; Tuscaloosa, AL: 1998. p. 508-573.
- Labov, W.; Ash, S.; Boberg, C. *Atlas of North American English*. Mouton de Gruyter; New York: 2005.
- Milroy, L. Social networks. In: Chambers, JK.; Trudgill, P.; Schilling-Estes, N., editors. *The handbook of language variation and change*. Blackwell; Malden, MA: 2002. p. 549-572.
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*. 1990; 47:379–390. [PubMed: 2345691]
- Nagy, N.; Zhang, X.; Nagy, G.; Schneider, EW. A quantitative categorization of phonemic dialect features in context. In: Dey, AEA., editor. *Context 2005*. Springer; Berlin: 2005. p. 326-338.
- Nerbonne J, Heeringa W. Computational comparison and classification of dialects. *Dialectologia et Geolinguistica*. 2001; 9:69–83.
- Niedzielski N. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*. 1999; 18:62–85.
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychological Science*. 1994; 5:42–46.
- Pisoni, DB. Some thoughts on “normalization” in speech perception. In: Johnson, K.; Mullennix, JW., editors. *Talker variability in speech processing*. Academic Press; San Diego: 1997. p. 9-32.
- Rakerd, B.; Plichta, B. More on perceptions of /a/ fronting; Paper Presented at New Ways of Analyzing Variation 32; Philadelphia, PA. October 9–12; 2003.
- Sattath S, Tversky A. Additive similarity trees. *Psychometrika*. 1977; 42:319–345.
- Tamasi, SL. *Cognitive patterns of linguistic perceptions*. University of Georgia; Athens: 2003. Doctoral dissertation

- Thomas, ER. An acoustic analysis of vowel variation in New World English. Duke University Press; Durham, NC: 2001.
- Thomas ER. Sociophonetic applications of speech perception experiments. *American Speech*. 2002; 77:115–147.
- Tice, R.; Carrell, T. Level16 (Version 2.0.3) [Computer Software]. University of Nebraska; Lincoln, NE: 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.
- Williams, A.; Garrett, P.; Coupland, N. Dialect recognition. In: Preston, DR., editor. *Handbook of perceptual dialectology*. John Benjamins; Philadelphia: 1999. p. 345-358.

Fig. 1.
Clustering solution for the listeners in Experiment 1.

Fig. 2.
Multidimensional scaling solution for the listeners in Experiment 1.

Fig. 3. Clustering solutions for the mobile North, mobile Midland, non-mobile North, and non-mobile Midland listener groups in Experiment 2.

Fig. 4. Multidimensional scaling solution for the listeners in Experiment 2 (Dimensions 1 and 2). The filled symbols represent male talkers and the open symbols represent female talkers. For the West, plain X's indicate females and boxed X's indicate males.

Fig. 5. Multidimensional scaling solution for the listeners in Experiment 2 (Dimensions 1 and 3). The filled symbols represent male talkers and the open symbols represent female talkers. For the West, plain X's indicate females and boxed X's indicate males.

Table 1

Characteristic properties of the six talker dialects in Experiment 1

Talker dialect	Characteristic acoustic-phonetic properties
New England	r-lessness in <i>dark</i> , /æ/ backing in <i>rag</i>
North	Centralized /ow/ offglides in <i>don't</i> , monophthongal /æ/ in <i>rag</i>
North Midland	None, among those analyzed
South Midland	/u/ fronting in <i>suit</i> , /ow/ backing in <i>don't</i>
South	<i>Greasy~greazy</i> alternation
West	None, among those analyzed

Table 2

Descriptive statistics on the number of talker groups produced by each listener group in Experiment 2

	Number of talker groups			
	Mean	Minimum	Maximum	Median
Mobile North	9.73	3	23	8
Mobile Midland	9.23	4	19	9
Non-mobile North	7.41	3	14	7
Non-mobile Midland	7.52	3	15	7
Overall	8.48	3	23	8

Table 3

Descriptive statistics on the number of talkers per group produced by each listener group in Experiment 2

	Number of talkers per group			
	Mean	Minimum	Maximum	Median
Mobile North	6.79	1	36	3
Mobile Midland	6.05	1	31	4
Non-mobile North	7.64	1	38	4
Non-mobile Midland	7.86	1	38	5
Overall	7.08	1	38	4

Table 4

Mean percent correct and error scores for each of the four listener groups (standard deviations are shown in parentheses)

	Percent correct	Percent errors
Mobile North	26 (17)	19 (16)
Mobile Midland	27 (14)	15 (10)
Non-mobile North	32 (16)	22 (13)
Non-mobile Midland	29 (18)	19 (15)
Overall	28 (16)	19 (13)

Table 5

Normalized subject weights for each of the four listener groups for each of the three dimensions in the INDSCAL analysis

	Dimension 1 (markedness)	Dimension 2 (gender)	Dimension 3 (geography)
Mobile North	0.4076	0.3093	0.2832
Mobile Midland	0.3804	0.3429	0.2767
Non-mobile North	0.4138	0.2789	0.3073
Non-mobile Midland	0.4029	0.3184	0.2788
Overall	0.4012	0.3123	0.2865