



Published in final edited form as:

*J Phon.* 2009 October 1; 37(4): 436–451. doi:10.1016/j.wocn.2009.07.004.

## Free classification of American English dialects by native and non-native listeners

**Cynthia G. Clopper** and

Ohio State University, Columbus OH 43210, USA

**Ann R. Bradlow**

Northwestern University, Evanston IL 60208, USA

### Abstract

Most second language acquisition research focuses on linguistic structures, and less research has examined the acquisition of sociolinguistic patterns. The current study explored the perceptual classification of regional dialects of American English by native and non-native listeners using a free classification task. Results revealed similar classification strategies for the native and non-native listeners. However, the native listeners were more accurate overall than the non-native listeners. In addition, the non-native listeners were less able to make use of constellations of cues to accurately classify the talkers by dialect. However, the non-native listeners were able to attend to cues that were either phonologically or sociolinguistically relevant in their native language. These results suggest that non-native listeners can use information in the speech signal to classify talkers by regional dialect, but that their lack of signal-independent cultural knowledge about variation in the second language leads to less accurate classification performance.

### Keywords

second language acquisition; dialect classification

### 1.0 Introduction

Learning the sound system of a second language involves the acquisition of a new phonological system, including new phoneme categories, new phonological rules or constraints, and new prosodic structures, as well as the acquisition of a new social indexical system, including phonological and phonetic markers for age, gender, and socioeconomic status. The acquisition of sociolinguistic knowledge in a second language is important not only for developing appropriate social communication skills in a second culture, but also has implications for speech processing. Native listeners exhibit processing benefits for familiar and standard varieties of their first language in a range of tasks, including sentence and word recognition (Clopper & Bradlow, 2008; Labov & Ash, 1997) and lexical decision (Floccia, Goslin, Girard, & Konopczynski, 2006). Knowledge about dialect variation<sup>1</sup>, therefore, may also have implications for online speech processing for non-native listeners.

---

Corresponding author: Cynthia G. Clopper, Department of Linguistics, Ohio State University, 1712 Neil Avenue, Columbus OH 43210, USA, clopper.1@osu.edu, Phone: 1 614 292 8235, Fax: 1 614 292 8833.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Most previous research on second language acquisition has focused on the linguistic properties, rather than the indexical properties, of the speech signal, and models of second language speech perception, such as Flege's (1995) Speech Learning Model and Best's (1995) Perceptual Assimilation Model, were developed to account for the perception of phonetic and phonological category differences across languages, rather than the perception of social variation in the second language. However, a small number of studies have explicitly examined the perception of sociolinguistic variation by second language learners. In one early study, Eisenstein (1982) conducted a perceptual dialect discrimination task with native and non-native English listeners in New York City. The non-native listeners included beginning, intermediate, and advanced learners of English. The participants were asked to discriminate between General American English, New York English, African American Vernacular English, Irish English, and Hawaiian Pidgin English. While dialect discrimination performance was good for all of the non-native listeners, the beginning and intermediate English learners performed more poorly overall than the most advanced English learners. However, the most advanced learners performed as well as the native listeners, suggesting that linguistic and cultural experience improves the perception of social indexical information in a second language.

Stephan (1997) explored the perception of world varieties of English by native German learners of English in an open-set dialect identification task. The varieties that the listeners were asked to identify included Northern and Southern British English, Cockney, Welsh English, Scottish English, Southern American English, Australian English, New Zealand English, South African English, West African English, and Indian English. The responses were scored generously as correct if the listener identified the correct region of the world. Identification accuracy differed considerably across the different varieties: Southern American English was correctly identified by 46% of the listeners, but South African English was correctly identified by only 3% of the listeners. The average accuracy across all of the varieties included in the experiment was 23%. Stephan (1997) attributed the differences in identification accuracy across the eleven varieties to differences in exposure. His informal survey of German EFL textbooks and German university course offerings in language variation in English suggested that German students have very little exposure to world varieties of English and that most of their experience is with standard British or American varieties.

Stephan (1997) did not directly compare his results to dialect identification performance by native listeners. However, Bayard, Weatherall, Gallois, and Pittam (2001) reported that New Zealand and Australian listeners could accurately identify New Zealand, Australian, North American, and European varieties of English in a forced-choice categorization task with approximately 84% accuracy. The New Zealand listeners correctly categorized the New Zealand, Australian, and American talkers with 85%, 57%, and 66% accuracy, respectively, although the relatively low accuracy for the Australian talkers was due to 49% mis-categorizations of one Australian talker as New Zealand. The Australian listeners correctly categorized the New Zealand, Australian, and American talkers with 83%, 84%, and 77% accuracy, respectively. American English listeners performed the same task with only 41% accuracy overall, correctly categorizing the New Zealand, Australian, and American talkers with 15%, 17%, and 87% accuracy, respectively. Bayard et al. (2001) attributed these differences in classification accuracy between the listener groups to asymmetric media exposure: American media is widespread in New Zealand and Australia, but Americans have little access to New Zealand or Australian media. Thus, American English is much more familiar to New Zealand and Australian listeners than New Zealand and Australian English are to American listeners. The differences in perceptual dialect classification accuracy between

---

<sup>1</sup>The term "dialect" is used here to describe variation among native speakers of a given language. While the focus of this study is phonological variation, the term "accent" is avoided to reduce potential ambiguity between native variation and non-native or foreign accent.

the listener groups in these tasks provide additional support for Stephan's (1997) suggestion that dialect classification performance is affected by linguistic exposure; the native listeners outperformed the non-native listeners and the New Zealand and Australian listeners outperformed the American listeners. However, given the wide range of varieties examined in Stephan's (1997) identification task and the large difference in performance among the native listeners in Bayard et al.'s (2001) categorization task, the overall difference in dialect classification accuracy by native and non-native listeners in these two studies should be interpreted with caution.

The role of linguistic experience has also been examined in studies of dialect intelligibility. Fox and McGory (2007) found that both native English listeners and Japanese learners of English performed better on a vowel identification task for General American vowels than for Southern American vowels. However, whereas native English listeners in Ohio performed better than native English listeners in Alabama on the General American vowels, Japanese learners of English did not show an intelligibility preference for the local English variety in either Ohio or Alabama. Eisenstein and Verdi (1985) also examined the effects of dialect familiarity on speech intelligibility performance for non-native listeners. For working class English learners in New York, African American Vernacular English was significantly less intelligible than either General American or New York English, despite the listeners' daily exposure to all three varieties. Thus, whereas native listeners show a processing benefit for both standard and local, familiar varieties (Clopper & Bradlow, 2008; Floccia et al., 2006; Labov & Ash, 1997), non-native listeners exhibit a processing benefit for standard dialects, but for only some local dialects.

Research on sociolinguistic attitudes suggests that second language learners also develop some native-like social stereotypes associated with different dialects of the second language. For example, Eisenstein (1982) found that both native and non-native listeners' judgments of level of education and socioeconomic status were significantly lower for African American Vernacular English and New York English than for General American English. As in her dialect discrimination task, however, Eisenstein (1982) observed that the more advanced learners' attitude judgments aligned more closely than the less advanced learners' judgments with those of native speakers. Similarly, Alford and Strother (1990) reported that both native and non-native English listeners judged Southern and Midwestern American English more positively than New York English in a matched-guise attitude judgment task. However, the non-native listeners consistently assigned lower overall ratings to the female talkers than the male talkers, unlike the native listeners who rated the male and female talkers within each dialect equally. Alford and Strother (1990) attributed this difference between the native and non-native listeners to cultural norms about gender that were shared by native listeners but that had not been acquired by the non-native listeners.

Taken together, research on the perception of sociolinguistic variation by non-native listeners suggests that non-native listeners can use variability in the speech signal to make judgments about dialect differences. However, non-native listeners tend to exhibit lower accuracy scores than native listeners in explicit dialect discrimination and identification tasks and to benefit less from dialect familiarity than native listeners in speech intelligibility tasks, suggesting that native listeners may also draw on cultural knowledge or stereotypes that non-native listeners have not fully acquired. This interpretation of these results is further supported by the results of attitude judgment tasks that reveal that explicit dialect attitudes are more native-like for more advanced learners of English than for beginning learners.

The task of explicit perceptual dialect classification and discrimination by non-native listeners involves two separate skills. First, the listeners must develop adequate sensitivity to sublexical and subphonemic variability in the acoustic signal to be able to reliably distinguish between

different varieties of the second language. For example, a non-native speaker of English must be able to distinguish monophthongal [ a : ] from diphthongal [ a j ] or [grisi] from [grizi] as alternative pronunciations of *greasy* to differentiate Southern American English from other varieties of American English. Second, the listeners must learn how the acoustic-phonetic properties of the signal combine to create constellations of variables that together index the different dialects. That is, any individual Southern talker may produce / a j / monophthongization, [grizi] for *greasy*, both, or neither. The non-native speaker of English must therefore learn, independent of the specific properties of any given Southerner's speech, which linguistic variants are associated with which dialect categories.

The purpose of the current study was to examine dialect classification performance by native and non-native listeners. To reduce the need for the non-native listeners to accurately identify the varieties of English presented in the experiment, a free classification task was used. The free classification task required the listeners to sort a set of talkers by regional dialect of American English, but it did not require them to provide category labels for their groups. While the target dialect categories were not explicitly provided to the listeners, the task could not be successfully completed without some notion of indexical categories (e.g., regional dialect) and how they are marked linguistically (e.g., by / a j / monophthongization). That is, the listeners were asked to judge the similarity of the dialects of the talkers and not the similarity of the voices of the talkers, and therefore, needed to ignore sources of talker-specific variability, such as overall pitch, in making their dialect classification judgments.

To reduce the effects of intelligibility on performance, the stimulus materials consisted of a single sentence produced by each of the different talkers. The use of a single sentence also allowed us to investigate the properties of the speech signal that the native and non-native listeners were using to make their classification judgments. In particular, we could explore how individual acoustic-phonetic variables, as well as constellations of two or more variables, contributed to judgments of dialect similarity. The non-native participants in this study had relatively little experience in the United States and were therefore expected to have less of the cultural knowledge about dialect variation that is shared by native speakers of American English. Thus, we were able to obtain a more fine-grained test of the hypothesis that both sensitivity to subphonemic variability, such as systematic vowel shifts in American English, and cultural knowledge about dialect-specific variation in the second language, such as how variants combine to create constellations of features that indicate dialect affiliation, contribute to differences in dialect classification performance between native and non-native listeners.

## 2.0 Experiment 1

### 2.1 Listeners

Forty-seven native speakers of American English and 36 non-native speakers of English were recruited from the Northwestern University community to participate as listeners in Experiment 1. Data from eleven of the native listeners were excluded from the analysis because the listeners were bilingual (N=8), reported a history of a hearing or speech disorder (N=2), or due to experimenter error in data collection (N=1). The remaining 36 monolingual native listeners (16 male, 20 female) ranged in age from 18-22 years old and received partial course credit in an introductory linguistics course for their participation in Experiment 1 and another, unrelated experiment. The native listeners represented a number of different regional dialects of American English, including Northern (N=21), Southern (N=1), Mid-Atlantic (N=1), and General American (N=10). The remaining three native listeners had lived in multiple different dialect regions before the age of 18.

The non-native listeners (17 male, 19 female) ranged in age from 16-32 years old and received \$8 for their participation in Experiment 1 and the same, unrelated experiment as the native

listeners. The non-native listeners represented a number of different first languages, including French (N=1), German (N=1), Gikuyu (N=1), Gujarati (N=1), Hindi (N=1), Italian (N=3), Korean (N=2), Mandarin (N=23), Tamil (N=2), and Telugu (N=1). Most (N=30) of the non-native listeners had spent less than one month in the United States at the time of the experiment. The remaining non-native listeners had spent either 1-2 months (N=3) or 2-3 years (N=3) in the United States at the time of the experiment. The English proficiency of the participants varied somewhat, but all were relatively proficient with written English as demonstrated by their Test of English as a Foreign Language (TOEFL) scores, which ranged from 600-673, with a mean of 634. Measures of proficiency in spoken English were not available for the non-native participants. None of the non-native listeners reported a history of a hearing or speech disorder.

Given the variation in the native languages of the non-native listeners, the non-native listeners were divided into two groups for analysis. The 13 listeners whose native language was not Mandarin made up the heterogeneous non-native listener group (Fennell, Byers-Heilein, & Werker, 2007). The 23 native Mandarin listeners comprised the Mandarin listener group. The separate analysis of the Mandarin listeners permitted a more concrete interpretation of the relationship between the native language of the non-native listeners, the acoustic-phonetic properties of the speech signal, and the perceptual dialect classification judgments.

## 2.2 Talkers

Twenty white male talkers were selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986). Five talkers were selected from each of four dialect regions in the United States (New England, North, Midland, and South). All of the selected talkers were in their 20s at the time of recording. The talkers were selected by the first author as good representatives of the target dialect regions, based on the phonetic characteristics of their speech. While the dialects have undergone additional changes since the TIMIT corpus was recorded, previous research with these talkers has revealed that college-aged native listeners can accurately classify them by regional dialect (Clopper & Pisoni, 2004, 2007).

## 2.3 Stimulus Materials

The TIMIT corpus includes recordings of 630 talkers reading 10 different sentences each. Two of the sentences were read by all of the talkers included in the corpus, and each of the remaining eight sentences were read by only a small subset (1-3) of the talkers. Thus, to be able use the same sentence for all of the talkers and avoid potential effects of intelligibility on dialect classification performance, we were limited to the two sentences read by all of the talkers in the corpus. Fortunately, the two sentences read by all of the talkers were written to elicit dialect-specific variation in American English (Fisher et al., 1986). The first sentence, “She had your dark suit in greasy wash water all year” produced by each of the 20 selected talkers was used in Experiment 1.

Previous acoustic analyses of the sentence produced by a larger set of male talkers, including those selected for the current study, revealed significant main effects of dialect on production (Clopper & Pisoni, 2004). Five phonetic properties were assessed for each talker by Clopper and Pisoni (2004): r-lessness in *dark*, / ʊ / backness in *suit*, fricative voicing in *greasy*, fricative duration in *greasy*, and r-lessness in *wash*. A series of multiple logistic regressions (Paolillo, 2002) was conducted to examine the relationship between these five phonetic properties and the actual dialect affiliation of the subset of 20 talkers used in the current study. For each dialect (New England, North, Midland, and South), the acoustic measures corresponding to the phonetic properties were entered as continuous independent variables and the talkers' actual dialect affiliation was the binary dependent variable. For example, in the New England analysis,

the five New England talkers were coded as 1 and the other 15 talkers were coded as 0. Significant pairwise correlations between the acoustic measures were observed for r-lessness in *dark* and /u/ backness in *suit* ( $r^2 = -0.34, p = .007$ ) and fricative voicing and duration in *greasy* ( $r^2 = -0.66, p < .001$ ). However, an inspection of the variance inflation factors in the logistic regression analyses revealed acceptable levels of collinearity (all VIF < 3).

A summary of the observed significant dialect differences in the logistic regression analyses is shown in Table 1. R-lessness in *dark* was a significant predictor of New England dialect affiliation and voicing of the fricative in *greasy* (i.e., [grizi] for /grisi/) was a significant predictor of Southern dialect affiliation. Despite these significant regression analyses, however, the talkers within each dialect exhibited variability in the extent to which they produced the dialect-specific variants. For example, the logistic regression model for the New England talkers that included the significant r-lessness variable correctly classified only 70% of the talkers into New England and non-New England dialects. The model for the Southern talkers that included voicing of the fricative in *greasy* correctly classified 90% of the talkers into Southern and non-Southern dialects, but the two talkers that were incorrectly classified were both Southerners that were misclassified as non-Southerners. Thus, the talkers in the current study exhibited variability both within and across the dialects. In addition, the talkers may have produced other dialect-specific properties that were not included in the acoustic analysis of the stimulus materials.

For the Mandarin listeners, r-lessness in *dark* and *wash* is expected to be salient because post-vocalic rhoticization serves to distinguish Shanghai from Beijing varieties of Mandarin. Rhotic codas are not observed in Shanghai, whereas they are a characteristic property of the Beijing dialect (Duanmu, 2000). Voicing of the fricative in *greasy* may also be salient for Mandarin listeners. Standard Mandarin does not have phonologically voiced fricatives, but it does have an aspiration distinction in affricates that is similar to the aspiration distinction in English stops, and the unaspirated Mandarin affricates can be phonetically voiced. In addition, the retroflex liquid /r/ is realized in some dialects of Mandarin as /z/ (Duanmu, 2000). Variation in the fronting of /u/ may be salient for Mandarin listeners if some of the variants assimilate to the Mandarin /u/ category and others assimilate to the Mandarin /y/ category (Best, 1995). The Mandarin listeners were not asked to identify the variety of Mandarin that they speak. Thus, the salience of these phonetic properties may vary across listeners, depending on their own native variety.

The stimulus materials were converted to digital movies for presentation to the listeners. Audio-visual stimulus materials were necessary to provide the listeners with visual icons linked to the auditory stimulus materials that could be moved around the screen. The visual track of each movie was a static image of a black rectangle with the talker's initials in white text. The static image allowed the same visual representation of the stimulus materials to be presented before, during, and after the audio track was played. The audio track of each movie was the sound file containing the talker's production of the target sentence. Prior to producing the movies, the audio stimulus files were equated for overall RMS amplitude. The movies were rendered at 16 frames per second with an audio sampling rate of 22050Hz and 16-bit resolution.

## 2.4 Procedure

Listeners were seated in front of personal computers equipped with headphones and a mouse in a sound-attenuated booth. The stimulus materials were presented in a single Powerpoint slide as shown in Figure 1 (see also Clopper, 2008). The 20 stimulus movies were presented to the left of a 16 × 16 grid. The target sentence was printed at the top of the slide. Participants could listen to the stimulus materials by double-clicking on them with the mouse. The stimulus materials were presented to the listeners at a comfortable listening level using the volume

control settings on the computers. The listeners could move the stimulus items around the screen by dragging them with the mouse.

The participants were instructed to listen to each of the talkers and then group them on the grid based on regional dialect. They were asked to put all of the talkers from the same part of the country in a group together. They could 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. The participants were able to listen to the talkers in any order and could listen to and move the talkers around the grid as many times as they wanted until they were satisfied with their solution.

## 2.5 Results

A summary of the listeners' classification strategy is shown in Table 2. The native, the heterogeneous non-native, and the Mandarin listeners all made approximately 6 groups of talkers with an average of 3-4 talkers per group. A one-way ANOVA confirmed that the difference in number of talker groups produced by the three groups of listeners was not significant ( $F(2, 71) = 0.5$ , n.s.).

Three measures of accuracy were calculated to assess classification performance. First, the number of correct talker pairings out of the total possible number of correct talker pairings (percent correct pairings) was calculated for each listener. Talker pairings were scored as correct if two talkers from the same dialect were put in a group together. The percent correct pairings score is similar to "hits" in a signal detection theory analysis (Macmillan, 1993). Second, the number of pairwise talker errors out of the total possible number of incorrect pairings (percent pairwise errors) was calculated for each listener. Talker pairings were scored as errors if two talkers from different dialects were put in a group together. The percent pairwise error score is similar to "false alarms" in a signal detection theory analysis. Since these two measures are not corrected for the number of talker groups that the individual listeners produced, the percent pairwise errors were subtracted from the percent correct pairings for each listener to obtain a difference score similar to  $d$ -prime (hits – false alarms) in a signal detection theory analysis.

As shown in Table 2, the native listeners produced a higher mean percent correct pairings, a lower mean percent pairwise errors, and a larger mean difference score than the heterogeneous non-native and Mandarin listeners. One-way ANOVAs confirmed significant effects of listener group for all three measures of accuracy ( $F(2, 71) = 9.9$ ,  $p < .001$  for percent correct pairings;  $F(2, 71) = 5.9$ ,  $p < .01$  for percent pairwise errors;  $F(2, 71) = 26.9$ ,  $p < .01$  for difference scores). Post-hoc Tukey tests confirmed significant differences between the native listeners and both groups of non-native listeners for the percent correct pairings and difference scores (all  $p < .01$ ) and between the native listeners and the Mandarin listeners for the percent pairwise errors ( $p < .01$ ). The difference between the native listeners and the heterogeneous non-native listeners for the percent pairwise errors was not significant. Thus, while the native and non-native listeners exhibited similar overall classification strategies, the native listeners were more accurate overall than the non-native listeners.

Clustering analyses were conducted to compare the perceptual dialect similarity structures for the native and non-native listeners. A  $20 \times 20$  talker similarity matrix was constructed for each listener in each listener group. When two talkers were grouped together, the value of the corresponding cell was set to 1. When two talkers were put in different groups, the value of the corresponding cell was set to 0. For each listener group, the individual listener matrices were summed, so that the group matrices reflected perceptual talker similarity in a range from 0-N, where 0 represented a pair of talkers who were never grouped together by any of the

listeners in the group (maximally dissimilar talkers) and N represented a pair of talkers who were grouped together by every listener in the group (maximally similar talkers).

The summed talker similarity matrices for the native, heterogeneous non-native, and Mandarin listener groups were submitted separately to the additive similarity tree analysis, ADDTREE (Corter, 1982). The resulting clustering solution for the native listeners is shown in Figure 2, the resulting clustering solution for the heterogeneous non-native listeners is shown in Figure 3, and the resulting clustering solution for the Mandarin listeners is shown in Figure 4. In these figures, perceptual similarity is represented by the length of the horizontal branches; the perceptual distance between any two talkers is the sum of the lengths of the fewest number of horizontal branches needed to connect them. Vertical distance is irrelevant. The lines indicating cluster divisions were added by hand to facilitate interpretation.

For the native listeners (Figure 2), three perceptual clusters emerged in the additive similarity tree analysis, accounting for 96% of the variance<sup>2</sup>. The top cluster includes all of the Northern and Midland talkers, the middle cluster includes all of the New England talkers, and the bottom cluster includes all of the Southern talkers. Thus, the native listeners exhibited a clear perceptual structure involving three dialects: New England, Southern, and Northern and Midland. A series of logistic multiple regression analyses was conducted to explore the relationship between the acoustic properties of the stimulus materials and the perceptual clusters obtained in the additive similarity tree analysis. For each cluster, the five acoustic measures shown in Table 1 were entered as independent variables and the talkers' membership in the cluster (1 for members and 0 for non-members) was the dependent variable. For example, talker Midland1 was coded as 1 for the analysis of the top cluster, but as 0 for the analyses of the middle and bottom clusters. The regression analyses revealed that voicelessness of the fricative in *greasy* was a significant predictor of membership in the top Northern and Midland cluster ( $\beta = -10.9, p = .006$ ), r-lessness in *dark* was a significant predictor of membership in the middle New England cluster ( $\beta = -.02, p = .005$ ), and voicing of the fricative in *greasy* was a significant predictor of membership in the bottom Southern cluster ( $\beta = 5.2, p = .002$ ).

For the heterogeneous non-native listeners (Figure 3), three perceptual clusters also emerged in the additive similarity tree analysis, accounting for 80% of the variance. The top cluster includes all of the New England talkers and one of the Southern talkers, the middle cluster includes all of the Midland and Northern talkers, and the bottom cluster includes the remaining four Southern talkers. Thus, the non-native listeners with mixed native languages perceived three similar dialect clusters to the native listeners, including New England, Southern, and Northern and Midland clusters, but the membership of individual talkers in each of the three clusters was not as clean. A series of logistic multiple regression analyses on cluster membership and the acoustic properties of the stimulus materials for the heterogeneous non-native listeners revealed that r-lessness in *dark* ( $\beta = -.02, p < .01$ ) was a significant predictor of membership in the top New England cluster, voicelessness of the fricative in *greasy* was a significant predictor of membership in the middle Midland and Northern cluster ( $\beta = -10.9, p < .01$ ), and voicing of the fricative in *greasy* was a significant predictor of membership in the bottom Southern cluster ( $\beta = 6.6, p < .001$ ). Thus, the non-native listeners were attending to the same acoustic properties as the native listeners in making their classification judgments.

For the Mandarin listeners (Figure 4), similar perceptual clusters emerged as for the other two listener groups, accounting for 75% of the variance. The top cluster includes all of the New England talkers and one Northern talker. The middle cluster includes three of the four Southern talkers that were in the heterogeneous non-native listeners' Southern cluster. The bottom cluster

<sup>2</sup>The variance accounted for by the ADDTREE analysis reflects the monotonic correlation between the input matrix distances and the output model distances.



includes all of the Midland talkers and the remaining Northern and Southern talkers. A series of logistic regression analyses on cluster membership and acoustic properties of the stimulus materials revealed that for the Mandarin listeners, r-lessness in *dark* was a significant predictor of membership in the top New England cluster ( $\beta = -.02, p = .001$ ), voicing of the fricative in *greasy* was a significant predictor of membership in the middle Southern cluster ( $\beta = 49.2, p < .001$ ), and voicelessness of the fricative in *greasy* ( $\beta = -3438.1, p < .001$ ), r-fulness in *dark* ( $\beta = 1.5, p < .001$ ), and r-lessness in *wash* ( $\beta = -.77, p < .001$ ) were significant predictors of membership in the bottom Northern and Midland cluster.

These results suggest that the Mandarin listeners were attending to fricative voicing and post-vocalic r-lessness to make their dialect classification judgments. Given that unaspirated affricates in Mandarin are phonetically voiced, the Mandarin listeners may have been able to transfer their aspiration distinction in affricates to the voicing distinction in English fricatives to distinguish between [grizi] and [grisi]. Given the regional variation in the production of /r/ in Mandarin (in coda position and in alternation with /z/), the Mandarin listeners' use of r-lessness in *dark* and *wash* and of fricative voicing in *greasy* to classify the American talkers by dialect is also consistent with how they might use the same properties to classify Mandarin talkers by dialect.

With respect to the Mandarin listeners' split between the three Southern talkers in the middle Southern cluster and the two Southern talkers in the bottom Northern and Midland cluster, posthoc listening to the stimulus materials revealed that the Southern talkers included in the Southern cluster all produced *greasy* as [grizi], whereas the Southern talkers included in the Northern and Midland cluster both produced *greasy* as [grisi]. The Southern talker that appeared in the New England cluster for the heterogeneous non-native listeners was also one of the talkers who produced *greasy* as [grisi]. Thus, whereas the native listeners were able to identify all of the Southern talkers as belonging to the same dialect group, regardless of their production of *greasy*, the non-native listeners relied more heavily on the *greasy* ~ *greazy* alternation in making their classifications. This suggests that while all three groups of listeners could use acoustic-phonetic properties of the signal to make their classifications, the native listeners also benefited from signal-independent knowledge about the constellation of features that characterize the Southern dialect. That is, other acoustic-phonetic properties in the speech of some, but not all, of the Southern talkers (such as /u/ fronting in *suit* or /ae/ diphthongization in *had*) may have been used by the native listeners to correctly identify each of the Southern talkers as Southern. The native listeners were therefore able to classify all of the Southern talkers together, despite differences in the dialect-specific variants in the stimulus materials that led to non-significant results in the logistic regression analysis on dialect affiliation shown in Table 1.

## 2.6 Discussion

The native and non-native listeners exhibited similar classification strategies in the regional dialect free classification task. All three groups of listeners produced an average of six groups of talkers with 3-4 talkers per group. In addition, the perceptual dialect similarity spaces for the native and non-native listeners were similar. While the individual listeners produced an average of six groups of talkers, the clustering analyses on the aggregate data revealed three clusters (New England, Southern, and Northern and Midland) for each of the three groups of listeners. Finally, similar acoustic correlates to perceptual similarity were observed for all three groups, including voicing of the fricative in *greasy* and r-lessness in *dark*.

The non-native listeners were less accurate overall than the native listeners, however, as shown by the classification accuracy analysis, as well as by the model fits of the clustering analyses. In addition, the Mandarin listeners attended to some acoustic properties that the native listeners did not, including r-lessness in *wash*. The non-native listeners also were less able to combine

multiple independent dialect-specific properties to perceive the Southern talkers as a single dialect group, and relied more heavily than the native listeners on the *greasy* ~ *greazy* alternation to classify the Southern talkers. The analysis of the Mandarin listeners' performance allowed for a more concrete interpretation of the results of the logistic regression analyses. In particular, the Mandarin listeners' attention to fricative voicing and post-vocalic r-lessness may reflect the phoneme inventory and sociolinguistic patterns observed in Mandarin.

Taken together, these results suggest that non-native listeners can use acoustic properties of the signal to make classification judgments about regional dialect. However, in this experiment, the acoustic properties that emerged as significant predictors of performance were all related to consonants and, moreover, were categorical ([s] vs. [z] in *greasy*, presence or absence of [r] in *dark*). However, most descriptions of regional dialect variation in American English focus on subphonemic vowel variation (e.g., Labov, Ash, & Boberg, 2006), which may be more subtle and less accessible to non-native listeners than the categorical consonant variables in the stimulus materials in this experiment. In addition, this distinction between categorical and subphonemic phenomena for consonants and vowels in English may or may not map on to similar distinctions in the non-native listeners' native languages. Experiment 2 was therefore designed to replicate Experiment 1 with a different sentence that included predominantly subphonemic vowel variation across dialects to explore how well the findings from Experiment 1 would generalize to different stimulus materials.

### 3.0 Experiment 2

#### 3.1 Listeners

Forty-two native speakers of American English and 33 non-native speakers of English were recruited from the Northwestern University community to participate as listeners in Experiment 2. Data from 13 native listeners who were bilingual and from 2 native listeners who reported a history of a hearing or speech disorder were removed prior to analysis. Data from 1 non-native listener were also removed due to experimenter error. The remaining 27 monolingual native listeners (10 male, 17 female) ranged in age from 17-25 years old and received partial course credit in an introductory linguistics course for their participation in Experiment 2 and another, unrelated experiment. The native listeners represented a number of different regional dialects of American English, including Northern (N=4), Southern (N=2), Mid-Atlantic (N=5), and General American (N=9). The remaining seven native listeners had lived in more than one dialect region before the age of 18.

The remaining non-native listeners (18 male, 14 female) ranged in age from 21-39 years old and received \$8 for their participation in Experiment 2 and another, unrelated experiment. The non-native listeners represented a number of different first languages, including Indonesian (N=1), Japanese (N=3), Korean (N=6), Mandarin (N=15), Marathi (N=1), Portuguese (N=2), Spanish (N=2), and Turkish (N=1). The native language of the remaining non-native listener was not reported at the time of testing. Most (N=23) of the non-native listeners had spent less than one month in the United States at the time of the experiment. The remaining non-native listeners had spent either 1-2 months (N=1), 3-6 months (N=3), 1 year (N=1), or 3-4 years (N=4) in the United States at the time of the experiment. The English proficiency of the participants varied somewhat, but all were relatively proficient with written English as demonstrated by their TOEFL scores, which ranged from 436-673, with a mean of 632. None of the non-native listeners reported a history of a hearing or speech disorder. As in Experiment 1, the non-native listeners were divided into two groups for analysis: a heterogeneous non-native group (N=17) and a native Mandarin group (N=15).

### 3.2 Talkers

The same talkers were used in Experiment 2 as in Experiment 1.

### 3.3 Stimulus Materials

The stimulus materials for Experiment 2 were the second TIMIT sentence, “Don't ask me to carry an oily rag like that” produced by each of the 20 talkers. As in Experiment 1, the sentences were converted to digital movies for presentation to the listeners.

Previous acoustic analyses of the second TIMIT sentence produced by a larger set of male talkers, including those selected for the current study, revealed significant main effects of dialect on production (Clopper & Pisoni, 2004). Six phonetic properties were assessed for each talker by Clopper and Pisoni (2004): /*ow*/ backness in *don't*, /*ow*/ diphthongization in *don't*, /*oj*/ diphthongization in *oily*, /*ae*/ backness in *rag*, /*ae*/ diphthongization in *rag*, and /*aj*/ diphthongization in *like*. A series of multiple logistic regressions was conducted to examine the relationship between the acoustic measures corresponding to the phonetic properties and the actual dialect affiliation of the subset of 20 talkers used in the current study. None of the pairwise correlations between the acoustic measures were significant. A summary of the observed significant dialect differences in the logistic regression analyses is shown in Table 3. /*ae*/ backness in *rag* was a significant predictor of New England dialect affiliation and /*ae*/ frontness in *rag* was a significant predictor of Northern dialect affiliation. /*oj*/ diphthongization in *oily* was a significant predictor of Midland dialect affiliation and /*oj*/ monophthongization in *oily* was a significant predictor of Southern dialect affiliation. Despite these significant regression analyses, however, the talkers within each dialect exhibited some variability in the extent to which they produced the dialect-specific variants. The logistic regression model for the New England talkers that included /*ae*/ backness correctly classified 100% of the talkers into New England and non-New England dialects. However, the logistic regression models for the Northern talkers that included /*ae*/ frontness and for the Midland talkers that included /*oj*/ diphthongization correctly classified 90% of the talkers, and the model for the Southern talkers that included /*oj*/ monophthongization correctly classified only 80% of the talkers. Thus, the talkers in the current study exhibited variability both within and across the dialects. As in Experiment 1, the talkers may also have produced other dialect-specific properties that were not included in the acoustic analysis of the stimulus materials.

For the Mandarin listeners, diphthongization of /*ow*, *oj*, *aj*/ is expected to be salient because diphthongs involving high vowel offglides are contrastive with monophthongs in Mandarin (Duanmu, 2000). The diphthongization of /*ae*/ may be less salient, however, because Mandarin does not have any low falling diphthongs. The backness of /*ae*, *ow*/ may also not be salient for Mandarin listeners because the mid and low parts of the vowel space are sparsely populated. In addition, previous research has found that Mandarin listeners perform poorly in perceptual identification tasks with English /*ae*/ and /*ɛ*/ (Flege, Bohn, & Jang, 1997). As in Experiment 1, the Mandarin listeners were not asked to identify the variety of Mandarin that they speak, and the salience of these phonetic properties may vary across listeners, depending on their own native variety.

### 3.4 Procedure

The procedure was the same as in Experiment 1, except that the sentence “Don't ask me to carry an oily rag like that” was printed at the top of the Powerpoint slide.

### 3.5 Results

A summary of the listeners' classification strategy is shown in Table 4. The native, the heterogeneous non-native, and the Mandarin listeners made approximately 6 groups of talkers

with an average of 3-4 talkers per group. A one-way ANOVA confirmed that the differences in the number of talker groups produced by the three groups of listeners were not significant ( $F(2, 58) = 0.2, n.s.$ ).

As in Experiment 1, three measures of accuracy were calculated to assess classification performance. As shown in Table 4, the native listeners produced a higher mean percent correct pairings, a lower mean percent pairwise errors, and a larger mean difference score than the heterogeneous non-native and Mandarin listeners. One-way ANOVAs confirmed significant differences due to listener group for all three measures of accuracy ( $F(2, 58) = 14.1, p < .001$  for percent correct pairings;  $F(2, 58) = 6.4, p < .01$  for percent pairwise errors;  $F(2, 58) = 50.4, p < .001$  for difference scores). Post-hoc Tukey tests confirmed significant differences between the native listeners and both groups of non-native listeners for all three measures of accuracy (all  $p < .05$ ).

A series of two-way ANOVAs on the three measures of accuracy with experiment (first or second) and listener group (native, heterogeneous non-native, or Mandarin) as between-subject factors revealed significant main effects of listener group for all three accuracy measures ( $F(2, 124) = 23.4, p < .001$  for percent correct pairings;  $F(2, 124) = 11.5, p < .001$  for percent pairwise errors;  $F(2, 124) = 69.9, p < .001$  for difference scores). The main effect of experiment and the experiment  $\times$  listener group interaction were not significant for any of the three accuracy measures. Post-hoc Tukey tests confirmed significant differences between the native listeners and both groups of non-native listeners for all three measures of accuracy (all  $p < .01$ ). Thus, the native listeners were more accurate overall than the non-native listeners, and this finding was consistent across the two experiments.

Talker similarity matrices were calculated for each individual listener and for each listener group. The talker similarity matrices for the native, non-native, and Mandarin listeners were submitted separately to an ADDTREE analysis (Cortier, 1982). The resulting clustering solution for the native listeners is shown in Figure 5, the resulting clustering solution for the heterogeneous non-native listeners is shown in Figure 6, and the resulting clustering solution for the Mandarin listeners is shown in Figure 7. For the native listeners (Figure 5), three perceptual clusters emerged in the additive similarity tree analysis, accounting for 93% of the variance. The top cluster includes all of the New England talkers and two Midland talkers, the middle cluster includes all of the Northern talkers and one Midland talker, and the bottom cluster includes all of the Southern talkers and the remaining two Midland talkers. Thus, the native listeners exhibited three perceptual dialect clusters: New England, Northern, and Southern. A series of logistic multiple regression analyses on cluster membership and the acoustic properties of the stimulus materials was conducted using the measurements in Table 3 as independent variables. The analyses revealed that /æ/ backness in *rag* and /a j/ diphthongization in *like* were significant predictors of membership in the top New England cluster ( $\beta = -17.4, p < .001$  for /æ/ backness;  $\beta = .22, p < .001$  for /a j/ diphthongization), /æ/ fronting was a significant predictor of membership in the middle Northern cluster ( $\beta = -.01, p = .003$ ), and /a j/ monophthongization was a significant predictor of membership in the bottom Southern cluster ( $\beta = -.01, p = .026$ ).

For the heterogeneous non-native listeners (Figure 6), four perceptual clusters emerged in the additive similarity tree analysis, accounting for 79% of the variance. The top cluster includes three of the Northern talkers, one New England talker, and one Midland talker. The second cluster includes two New England talkers, one Northern talker, one Midland talker, and one Southern talker. The third cluster includes one New England talker, one Northern talker, one Midland talker, and two Southern talkers. The bottom cluster includes the remaining two Southern talkers, one New England talker, and two Midland talkers. These clusters roughly reflect three broad perceptual dialect categories: New England, Northern, and Southern, with

the Southern talkers split between the two bottom clusters. As in Experiment 1, the membership of individual talkers in each dialect cluster was less clean for the non-native listeners than the native listeners. A series of logistic multiple regression analyses on cluster membership and the acoustic properties of the stimulus materials revealed that /*oʊ*/ diphthongization ( $\beta = -.01$ ,  $p = .009$ ) and /*æ*/ monophthongization ( $\beta = -.03$ ,  $p = .001$ ) were significant predictors of membership in the top Northern cluster, and /*oʊ*/ backing ( $\beta = -.50$ ,  $p < .001$ ), /*æ*/ diphthongization ( $\beta = .48$ ,  $p < .001$ ), and /*oʊ*/ monophthongization ( $\beta = -.56$ ,  $p < .001$ ) were significant predictors of membership in the third Southern cluster. No significant acoustic properties emerged as predictors of membership in the second New England cluster or the bottom Southern cluster. Thus, in addition to the substantial overlap among the dialects within the four clusters, the heterogeneous non-native listeners were also not attending to any of the acoustic properties that the native listeners used in making their classifications.

The clustering solution for the Mandarin listeners (Figure 7) also reveals four perceptual dialect clusters, accounting for 74% of the variance. The top cluster includes four of the New England talkers, one Northern talker, and one Midland talker. The second cluster includes one Northern, one Midland, and one Southern talker. The third cluster includes the remaining three Northern talkers, one Midland, and one Southern talker. The bottom cluster includes the remaining three Southern talkers, two Midland talkers, and one New England talker. These clusters roughly reflect the three perceptual categories observed in the native and heterogeneous non-native listeners' solutions (New England, Northern, and Southern), plus a fourth cluster with one Northern, one Midland, and one Southern talker. As in the heterogeneous non-native listeners' solution, the distribution of talkers in each of the four clusters is less clean than what was observed for the native listeners. However, the Mandarin listeners' solution is cleaner than the heterogeneous non-native listeners' solution.

A series of logistic multiple regression analyses on cluster membership and the acoustic properties of the stimulus materials revealed that /*æ*/ backing ( $\beta = .01$ ,  $p = .03$ ) was a significant predictor of membership in the top New England cluster, and /*æ*/ fronting ( $\beta = -.49$ ,  $p < .001$ ) and /*ow*/ monophthongization ( $\beta = -.25$ ,  $p < .001$ ) were significant predictors of membership in the third Northern cluster. No significant acoustic properties emerged as predictors of membership in the bottom Southern cluster or the second cluster with one Northern, one Midland, and one Southern talker. Thus, the Mandarin listeners were attending to /*æ*/ fronting and /*ow*/ monophthongization in making their classification judgments. It is somewhat surprising that the Mandarin listeners were sensitive to variability in /*æ*/ fronting, given that the low front part of the Mandarin vowel space is sparsely populated and that previous research has found that Mandarin listeners perform poorly in perceptual identification tasks with English /*æ*/ and /*ɛ*/ (Flege, Bohn, & Jang, 1997). It is less surprising, however, that the Mandarin listeners relied on /*ow*/ monophthongization. The mid vowels /*ɪ*/ and /*ɛ*/ in Mandarin are phonologically contrastive with respect to diphthongization (Duanmu, 2000). Thus, similar to the *greasy-greazy* distinction in Experiment 1, the Mandarin listeners may have used /*ow*/ monophthongization to classify English talkers because diphthongization is also phonologically contrastive in their native language. However, unlike the native English listeners, the Mandarin listeners did not attend to /*aɪ*/ diphthongization in making their classification judgments. This finding is surprising given that /*aɪ*/ contrasts with /*a*/ in Mandarin (Duanmu, 2000) and that the Mandarin listeners were sensitive to diphthongization of another vowel (/*ow*/) in the same stimulus materials.

### 3.6 Discussion

The native and non-native listeners exhibited similar overall classification strategies in the free classification task. All three groups produced an average of 6 groups of talkers with 3-4 talkers per group. In addition, the perceptual dialect similarity spaces for the native and non-native

listeners were similar. The clustering analyses revealed New England, Southern, and Northern clusters for all three groups of listeners. The non-native listeners were less accurate overall than the native listeners, however, as indexed by the classification accuracy analysis, as well as by the model fits of the clustering analyses. In addition, the non-native listeners did not attend to all of the acoustic properties that were relevant for the native listeners, including / a j / monophthongization in *like*.

Unlike in Experiment 1, the results of the clustering analyses revealed differences in the perceptual dialect similarity structures between the native and non-native listeners. The clustering solutions for both groups of non-native listeners revealed four clusters of talkers, whereas the solution for the native listeners revealed three clusters. The regression analyses revealed that the heterogeneous non-native listener group did not attend to either of the cues that the native listeners did, whereas both the native listeners and the Mandarin listeners attended to / a e / fronting. In addition, the Mandarin listeners and the heterogeneous non-native listeners did not attend to any of the same cues in making their classification judgments.

Neither the heterogeneous non-native listeners nor the Mandarin listeners attended to / a j / diphthongization in classifying the talkers by dialect. One prediction that emerges from this finding is that / a j / diphthongization should be less salient than / a e / fronting for these non-native listeners. That is, the difference between [ a j ] and [ a ] should be less salient than the difference between [ a e ] and fronted [ a e ]. Additional research is needed to verify the relative perceptual salience of these types of subphonemic differences for native and non-native listeners, as well as how perceptual salience interacts with the first language of the non-native listeners.

#### 4.0 General Discussion

In both experiments, the three groups of listeners made approximately six groups of talkers with 3-4 talkers per group. However, the native listeners were significantly more accurate than the non-native listeners in both experiments. The difference in accuracy between the native and non-native listeners is consistent with Eisenstein's (1982) finding that beginning learners of English, who had been in New York City for an average of 7 months at the time of testing, performed more poorly on the perceptual dialect discrimination task than the more advanced learners. Most of the non-native listeners in the current study had been in the United States for less than one month at the time of testing and therefore had even less direct exposure to dialect variation in American English than Eisenstein's (1982) participants. However, information about experience with native speakers of American English was not obtained from the non-native participants in this study, and some may have been exposed to American English in the classroom through pedagogical materials or American instructors. In addition, some participants may have taken courses in dialect variation in American English prior to arriving in the United States. Thus, the results of the current study may provide additional support for Alford and Strother's (1990) intuition that both linguistic and cultural experience play an important role in developing accurate judgments about dialect variation in a second language. In particular, the non-native listeners in the current study were reasonably proficient in written English, but most had spent very little time in an English-speaking environment. Thus, they may have had access to the linguistic aspects of variability necessary to sort the talkers by dialect, but may not have had knowledge about the co-occurrence of variants within a dialect (such as / a j / monophthongization and / u / fronting in Southern American English) to help them group together talkers from the same dialect region with different constellations of variants in their speech. However, additional research is needed to examine how explicit instruction in dialect variation and/or exposure in the classroom to American English affects dialect classification performance by non-native listeners.

In the clustering analyses in both Experiments 1 and 2, the native solutions were cleaner than the non-native solutions, as indicated by the proportions of variance accounted for by the models. In addition, the solutions were cleaner for both the native and non-native listeners in Experiment 1 than Experiment 2. This finding is consistent with previous studies using these sentences for dialect classification among native listeners (e.g., Clopper & Pisoni, 2004); performance on the first TIMIT sentence is typically better than performance on the second TIMIT sentence. In addition, the clustering analyses revealed different perceptual similarity structures for the two sentences. In both experiments, the clustering solutions for all three groups of listeners revealed New England, Southern, and Northern clusters. In Experiment 1, the majority of the Midland talkers were clustered with the Northern talkers. In Experiment 2, however, the Midland talkers were more evenly distributed among the three clusters. This result is also consistent with previous dialect classification research using these sentences, which revealed greater perceptual similarity between Midland and Northern talkers for the first TIMIT sentence than the second TIMIT sentence (Clopper & Pisoni, 2004).

It is well established that perceptual sensitivity to specific phonemic and subphonemic differences is strongly affected by the relationship between the listener's native language and the second language (e.g., Best, 1995). These effects of native language on perceptual sensitivity were also observed in the current study; the perceptual similarity structures and acoustic-phonetic correlates of dialect similarity varied for the heterogeneous group of non-native listeners with mixed native languages and the listeners who shared a native language (Mandarin). For the native Mandarin listeners, two of the consonant phenomena (r-lessness in *dark* and r-lessness in *wash*) are also sociolinguistically relevant in Mandarin, and one of the consonant phenomena (fricative voicing) and one of the vowel phenomena (/ow/ monophthongization) are phonologically relevant in Mandarin. Thus, the differences in the cues that were attended to by the heterogeneous listeners and the Mandarin listeners may reflect the relationship between phonological and sociolinguistic patterns in the first and second languages. Additional research is needed to explore how phoneme inventory and sociolinguistic patterns in the native language affect the perception of sociolinguistic patterns in a second language.

Given that /æ/ is often one of the more difficult vowels for non-native learners of English to acquire (e.g., Bohn & Flege, 1997; Flege et al., 1997; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001), particularly in contrast to neighboring vowels such as /ɛ/ and /a/, the use of the variability in /æ/ to distinguish dialects was unexpected for the Mandarin listeners. Additional research is therefore also needed to explore the relationship between the perception of subphonemic variation in linguistic tasks, such as vowel identification and word recognition, and in sociolinguistic tasks, such as dialect classification, to account for this finding.

The lower accuracy scores and noisier clustering solutions for the non-native listeners may also reflect attention to acoustic-phonetic properties of the signal that were not good indicators of dialect affiliation, such as r-lessness in *wash* and /æ/ monophthongization in *rag*. This attention to inappropriate acoustic-phonetic cues suggests that the non-native listeners used information in the signal to classify the talkers by dialect, but they were less able to differentiate between reliable and unreliable cues to dialect affiliation than the native listeners. In addition, the non-native listeners did not attend to some appropriate cues to dialect affiliation, such as /aɪ/ diphthongization in *like*. Thus, their poorer overall performance may also reflect their inability to recognize constellations of cues that together indicate a given dialect, such as /ɔɪ/ monophthongization in *oily* and /aɪ/ monophthongization in *like* each independently signaling Southern dialect affiliation. That is, a given Southern talker may exhibit one, both, or neither of these two properties. However, native listeners may be able to use their knowledge of the constellations of cues that signal the Southern dialect to classify talkers together who do not necessarily exhibit overlapping variants in their speech. Non-native listeners, on the other

hand, may lack this signal-independent knowledge about the co-occurrence of variants in a given dialect of the second language and therefore, may not be able to ignore irrelevant variation in the speech signal in the perceptual dialect classification task. The results of the free classification task provide no evidence that the non-native listeners were basing their responses on overall voice similarity, but instead suggest that the listeners were relying on reasonable and reliable segmental properties for classifying the talkers by dialect. This attention to potentially relevant cues may reflect a universal familiarity with linguistic variation, due to experience with variability in the first language.

The results of this study suggest that indexical categories are acquired along with phonological categories in second language acquisition. The proficiency levels of the non-native listeners confirm their acquisition of some aspects of the English phonological system, and their ability to perform the dialect classification task with some success suggests that they have also acquired some knowledge about regional dialect categories in American English. Thus, second language learners can use information in the signal to make judgments about indexical categories, even with limited direct experience with dialect variation in the second language. However, learning which indexical categories are associated with different phonological variants and which variants make up the constellations that mark indexical categories for native listeners requires either greater proficiency or, more likely, more direct experience with variation in the second language.

Additional research is needed with more proficient and/or longer-term residents in a second language environment to explore the trajectory of development of signal-independent cultural knowledge about indexical variation in a second language.

## Acknowledgments

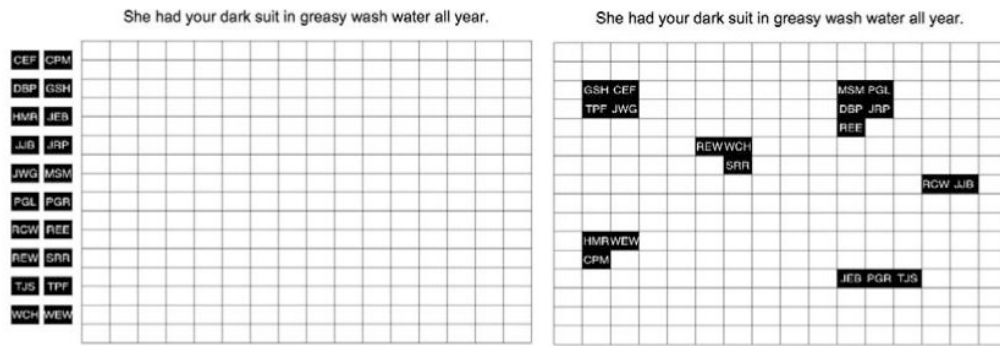
This work was supported by NIH F32 DC007237 and NIH R01 DC005794 to Northwestern University. Portions of this work were presented at the 10<sup>th</sup> Laboratory Phonology conference in 2006 and the 16<sup>th</sup> International Congress of Phonetic Sciences in 2007. The authors would like to thank Jennifer Alexander, Midam Kim, Kelsey Mok, Page Piccinini, Judy Song, Josh Viau, and Xiaoju Zheng for their assistance with data collection, and Fangfang Li for her assistance in interpreting the Mandarin listeners' data.

## Reference List

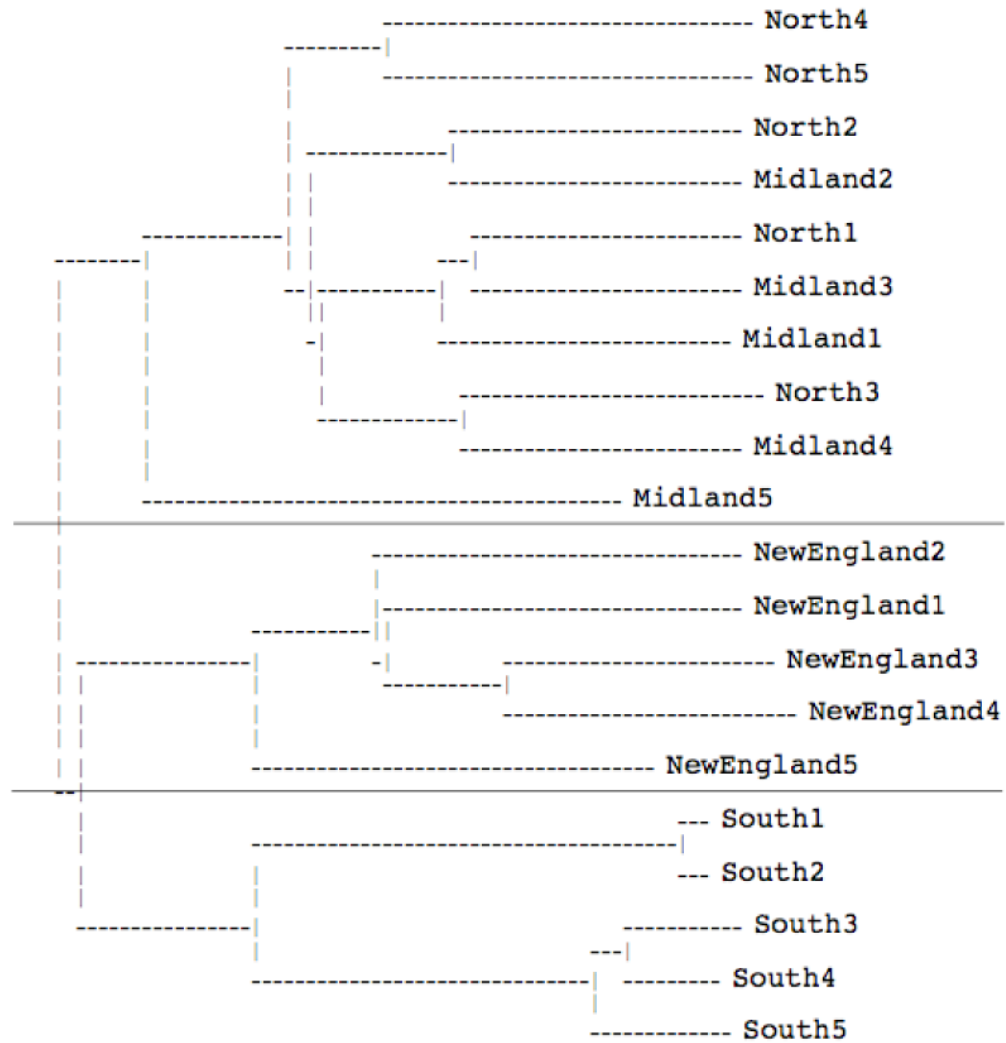
- Alford RL, Strother JB. Attitudes of native and nonnative speakers toward selected regional accents of U.S. English. *TESOL Quarterly* 1990;24:479–495.
- Bayard D, Weatherall A, Gallois C, Pittam J. Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics* 2001;5:22–49.
- Best, CT. A direct realist view on cross-language speech perception. In: Strange, W., editor. *Speech Perception and Linguistic Experience*. Timonium, MD: York Press; 1995. p. 171-204.
- Bohn, OS.; Flege, JE. Perception and production of a new vowel category by adult second language learners. In: James, A.; Leather, J., editors. *Second-Language Speech: Structure and Process*. Berlin: Mouton de Gruyter; 1997. p. 53-73.
- Clopper CG. Auditory free classification: Methods and analysis. *Behavior Research Methods* 2008;40:575–581. [PubMed: 18522069]
- Clopper CG, Bradlow AR. Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech* 2008;51:175–198. [PubMed: 19626923]
- Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 2004;32:111–140.
- Clopper CG, Pisoni DB. Free classification of regional dialects of American English. *Journal of Phonetics* 2007;35:421–438.
- 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.



- Duanmu, S. *The Phonology of Standard Chinese*. Oxford; Oxford University Press; 2000.
- Eisenstein M. A study of social variation in adult second language acquisition. *Language Learning* 1982;32:367–391.
- Eisenstein M, Verdi G. The intelligibility of social dialects for working-class adult learners of English. *Language Learning* 1985;35:287–298.
- Fennell CT, Byers-Heilein K, Werker JT. Using speech sounds to guide word learning: The case of bilingual infants. *Child Development* 2007;78:1510–1525. [PubMed: 17883445]
- Fisher WM, Doddington GR, Goudie-Marshall KM. The DARPA speech recognition research database: Specification and status. *Proceedings of the DARPA Speech Recognition Workshop* 1986:93–99.
- Flege, JE. Second-language speech learning: Theory, findings, and problems. In: Strange, W., editor. *Speech Perception and Linguistic Experience*. Timonium, MD: York Press; 1995. p. 233-277.
- Flege JE, Bohn OS, Jang S. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 1997;25:437–470.
- Floccia C, Goslin J, Girard F, Konopczynski G. Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance* 2006;32:1276–1293. [PubMed: 17002537]
- Fox, RA.; McGory, JT. Second language acquisition of a regional dialect of American English by native Japanese speakers. In: Bohn, OS.; Munro, MJ., editors. *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins; 2007. p. 117-134.
- 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. *The Atlas of North American English*. Berlin: Mouton de Gruyter; 2006.
- Macmillan, NA. Signal Detection Theory as data analysis method and psychological decision model. In: Keren, G.; Lewis, C., editors. *Methodological and Quantitative Issues in the Analysis of Psychological Data*. Vol. 1. Hillsdale, NJ: Erlbaum; 1993. p. 21-57.
- Paolillo, JC. *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford, CA: CSLI Publications; 2002.
- Stephan, C. The unknown Englishes? Testing German students' ability to identify varieties of English. In: Schneider, EW., editor. *Englishes around the World*. Amsterdam: John Benjamins; 1997. p. 93-108.
- Strange W, Akahane-Yamada R, Kubo R, Trent SA, Nishi K. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America* 2001;109:1691–1704. [PubMed: 11325137]

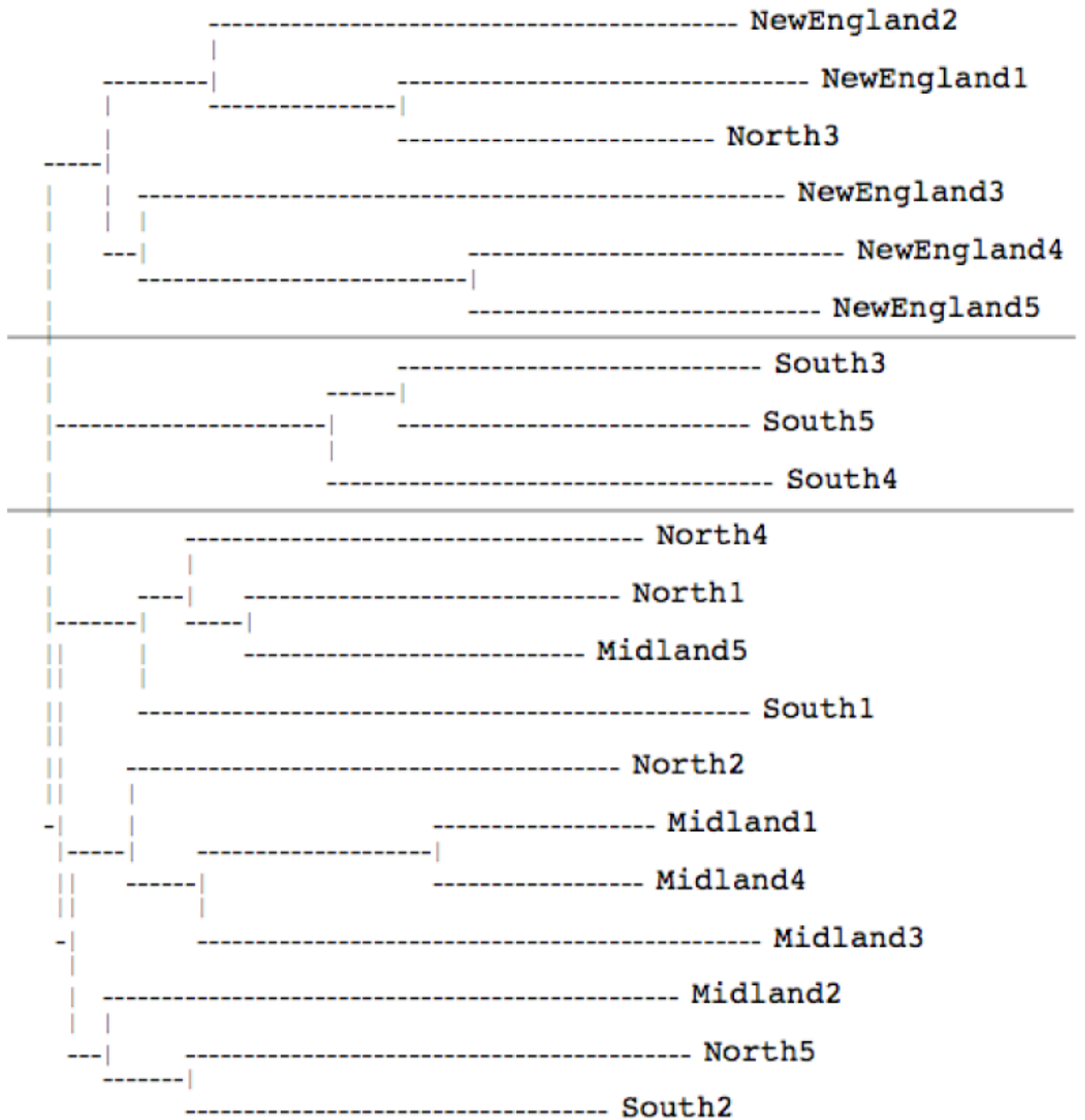


**Figure 1.** Stimulus presentation before (left) and after (right) the free classification task.

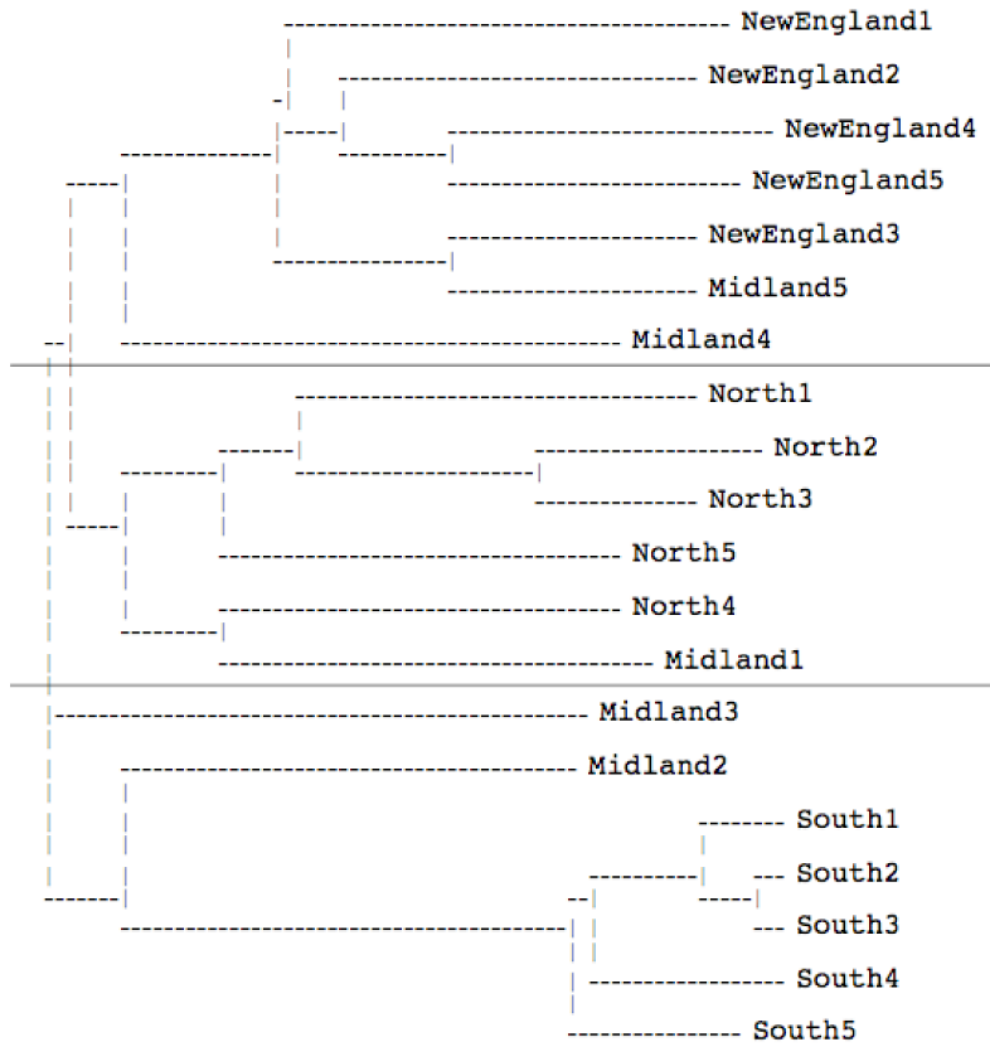


**Figure 2.**  
Clustering solution for the native listeners in Experiment 1.

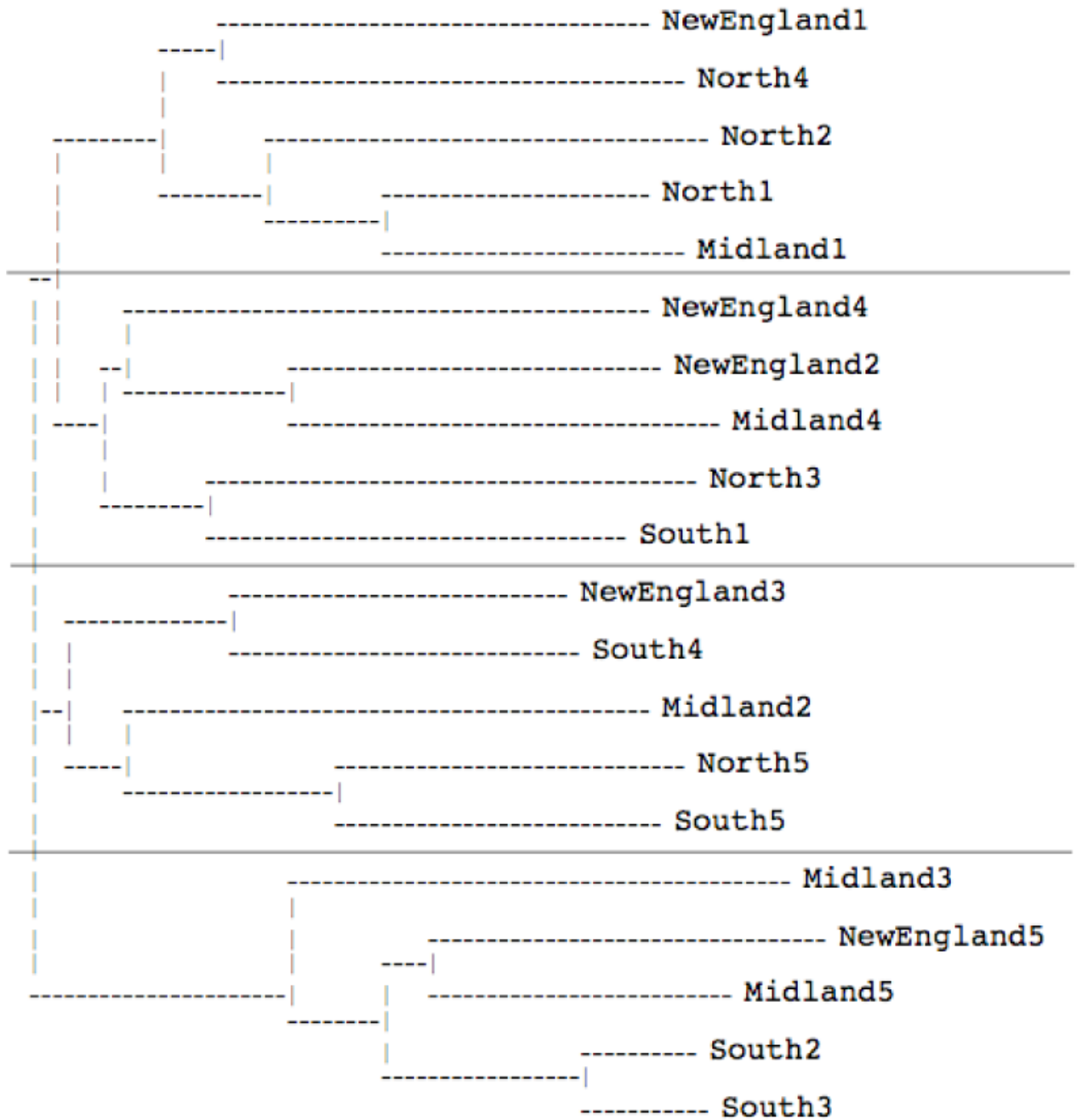




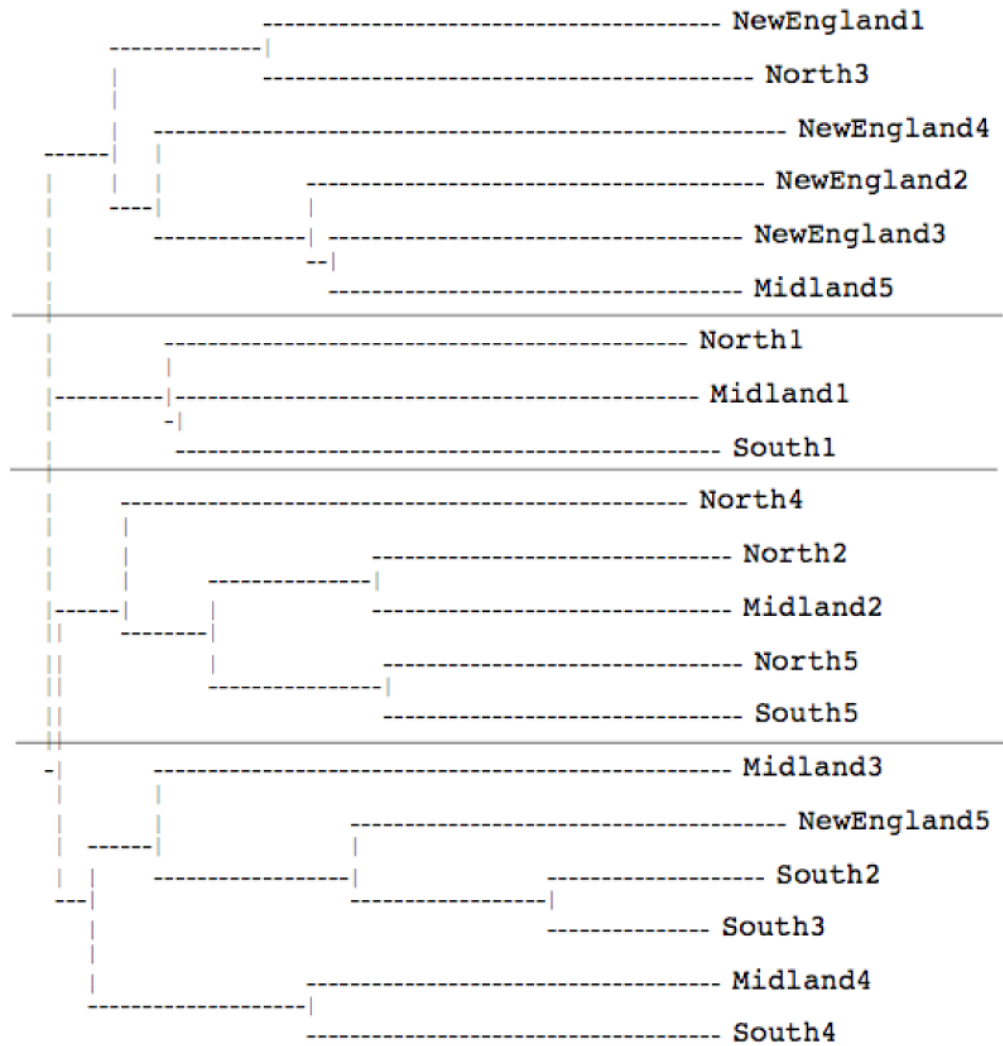
**Figure 4.**  
Clustering solution for the native Mandarin listeners in Experiment 1.



**Figure 5.** Clustering solution for the native listeners in Experiment 2.



**Figure 6.**  
Clustering solution for the heterogeneous non-native listeners in Experiment 2.



**Figure 7.** Clustering solution for the native Mandarin listeners in Experiment 2.



**Table 1**

Acoustic correlates of regional dialects of American English in the first TIMIT sentence.

<b>Word</b>	<b>Phonetic Property</b>	<b>Acoustic Measure</b>	<b>Correlated Dialect Affiliation</b>
dark	r-lessness	F3 midpoint – F3 offset	New England
suit	/u/ backness	F2 midpoint (normalized to F2 in 'year')	
greasy	fricative voicing	proportion of voicing	South
greasy	fricative duration	duration (normalized to word duration)	
wash	r-lessness	F3 midpoint	

**Table 2**

Summary of the classification strategy of the native and non-native listeners in Experiment 1.

	Native Listeners	Heterogeneous Non-native Listeners	Mandarin Listeners
<b>Number of Talker Groups</b>	6.11	6.62	6.04
<b>Talkers per Group</b>	3.53	3.25	3.69
<b>Percent Correct Pairings</b>	43	25	24
<b>Percent Pairwise Errors</b>	10	16	18
<b>Difference Score</b>	33	9	6
<b>(Correct – Errors)</b>			

**Table 3**

Acoustic correlates of regional dialects of American English in the second TIMIT sentence.

Word	Phonetic Property	Acoustic Measure	Correlated Dialect Affiliation
don't	/ow/ backness	F2 midpoint (normalized to F2 in 'year')	
don't	/ow/ diphthongization	F2 midpoint – F2 offset	
oily	/o j/ diphthongization /oj/ monophthongization	F2 offset – F2 midpoint	Midland South
rag	/ae/ backness /ae/ frontness	F2 midpoint (normalized to F2 in 'year')	New England North
rag	/ae/ diphthongization	F2 offset – F2 onset	
like	/a j/ diphthongization	F2 offset – F2 midpoint	

**Table 4**

Summary of the classification strategy of the native and non-native listeners in Experiment 2.

	Native Listeners	Heterogeneous Non-native Listeners	Mandarin Listeners
<b>Number of Talker Groups</b>	6.22	6.18	5.87
<b>Talkers per Group</b>	3.43	3.88	4.23
<b>Percent Correct Pairings</b>	44	21	24
<b>Percent Pairwise Errors</b>	9	18	17
<b>Difference Score</b>	35	3	7
<b>(Correct – Errors)</b>			