NATIVE AND NON-NATIVE PERCEPTUAL DIALECT SIMILARITY SPACES

Cynthia G. Clopper and Ann R. Bradlow

Ohio State University, Northwestern University
clopper.1@osu.edu, abradlow@northwestern.edu

ABSTRACT

The current study examined the role of native language on the perceptual similarity space of regional dialect variation. Native and non-native speakers of American English were asked to group a set of talkers by regional dialect in a free classification task. The two listener groups exhibited similar dialect classification strategies and perceptual similarity structures. However, the non-native listeners were less accurate overall than the native listeners and relied heavily on a few salient acoustic cues to make their classifications. These results suggest that non-native listeners can use lawful variation in the acoustic signal to make dialect classification judgments, but that cultural and linguistic familiarity also play a role in shaping perceptual dialect categories.

Keywords: dialect perception, free classification, non-native speech perception

1. INTRODUCTION

In acquiring the sound system of a second language, learners must master not only a new phonological system, including the segmental inventory, phonotactics, and prosody of the second language, but also new patterns of indexical variation, including regional, social, and ethnic dialects. Sociolinguistic variation provides essential information about the permissible range of variation for the second language segmental categories, as well as essential information about how social identity is conveyed in the second language. However, most second language acquisition research has focused on the former problem (e.g., [8]) and very little is known about non-native listeners' knowledge of second language sociolinguistic variation.

A range of explicit perception tasks assessing the classification of regional varieties of American English have shown that native listeners have explicit knowledge about sociolinguistic variation in their first language. For example, performance in forced-choice perceptual dialect categorization tasks is above chance and the patterns of errors produced by listeners in these tasks reflect regional phonological variation [2]. In addition, regional dialect similarity rating [1] and free classification [3] tasks have revealed three primary perceptual dialects of American English (Northeastern, Southern, and General American), along two perceptual dimensions of similarity: geography (North vs. South) and markedness (many vs. few characteristic properties). Non-native listeners have been shown to accurately identify world varieties of English in an open-set perceptual dialect identification task [7].

The goal of the current study was to explore the perceptual dialect similarity structure of non-native speakers of English who had only recently arrived in the United States. Specifically, we wanted to examine how perceptual dialect classification strategies and similarity spaces differed between native and non-native listeners. To reduce the role of differences in geographic and cultural knowledge between the native and non-native listeners, we used a perceptual free classification task [6] to assess dialect classification. The free classification paradigm provides participants with the opportunity to classify stimulus objects without the constraints of experimenter-defined category labels (e.g., *North*, *South*), a set number of categories, or specific dimensions of contrast.

2. METHODS

2.1. Listeners

Two groups of listeners participated in the free classification task. The first group included 36 non-native speakers of English (age 16-32 years old). The non-native listeners had a range of 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). In addition, the non-native listeners exhibited a range of proficiency levels in English, but all had TOEFL

scores greater than 600. Most (N=32) of the non-native listeners had been in the United States for less than two months at the time of the experiment.

The second group of listeners included 36 monolingual native speakers of American English (age 18-22 years old). The native listeners represented a range of regional dialects of American English, including Mid-Atlantic (N=1), Midland (N=6), Northern (N=21), Southern (N=1), and Western (N=4) varieties, as well as three listeners who had lived in more than one dialect region prior to age 18.

The native and non-native listeners were recruited from the Northwestern University community and were paid $8 for their participation. None of the listeners reported a history of a hearing or speech disorder. All of the listeners completed the free classification task after an unrelated experiment.

2.2. Talkers

Twenty white male talkers were selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus [5]. The talkers were 20-29 years old at the time of recording and represented four regional dialects of American English: New England (N=5), North, (N=5) Midland (N=5), and South (N=5).

2.3. Stimulus Materials

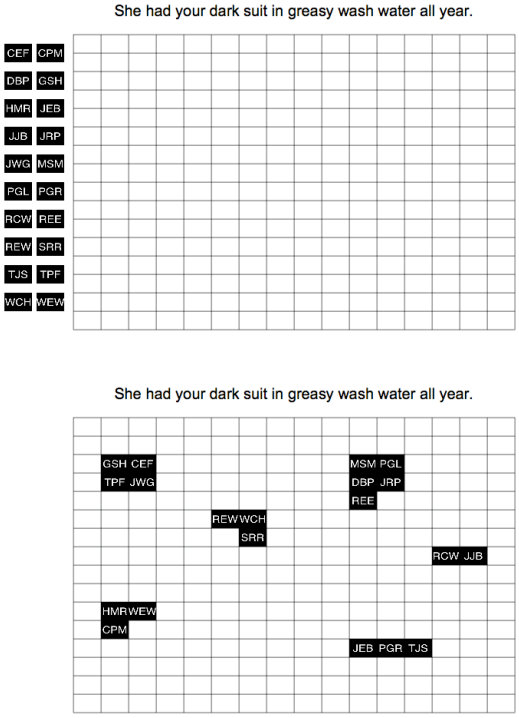
The stimulus materials consisted of a recording of each of the 20 talkers producing the sentence, "She had your dark suit in greasy wash water all year." This sentence was originally written to elicit regional dialect variation [5] and an acoustic analysis of dialect variables in the sentence confirmed that r-lessness in *dark* was significantly correlated with New England speech, s-voicing in *greasy* was significantly correlated with Southern speech, and Northern and Midland speech were not significantly correlated with any of the features examined [2].

2.4. Procedure

Participants were seated in a sound-attenuated booth in front of a PC equipped with headphones and a mouse. On the screen, the participants saw the sentence, "She had your dark suit in greasy wash water all year" printed above a 16 x 16 grid. To the left were 20 rectangles with three-letter sequences corresponding to the talkers' initials.

Each of the twenty rectangles was linked to a sound file produced by one of the talkers. The participants could listen to the stimulus materials by double-clicking on the rectangles and move the rectangles around the screen by dragging them with the mouse. The listeners were asked to group the talkers by regional dialect of American English. The task was unconstrained with respect to the number of groups, the number of talkers per group, or the number of times the stimulus items could be listened to or re-arranged. Figure 1 shows the stimulus presentation before (top) and after (bottom) the task.

Figure 1: Stimulus presentation before (top) and after (bottom) the task.



3. RESULTS

Two sets of analyses were conducted on the classification data. First, we examined the overall classification strategy of the two groups of listeners, including the number of groups of talkers produced and the overall accuracy of the classification. Second, we used a clustering analysis to explore the perceptual similarity spaces of the talkers for the native and non-native listeners.

A summary of the classification strategy of the two listener groups is shown in Table 1. Both the native and non-native listeners made approximately 6 groups of talkers, with an average of 3-4 talkers per group. A t-test confirmed that the

two listener groups did not differ in the number of talker groups they produced ($t(70)$=-.33, n.s.). However, out of the total possible number of correct and incorrect pairings, the non-native listeners made significantly fewer correct pairings ($t(70)$=4.47, p<.001) and significantly more pairwise errors ($t(70)$=-3.41, p=.001), respectively, than the native listeners. Thus, the non-native listeners performed less accurately than the native listeners, despite a similar classification strategy.

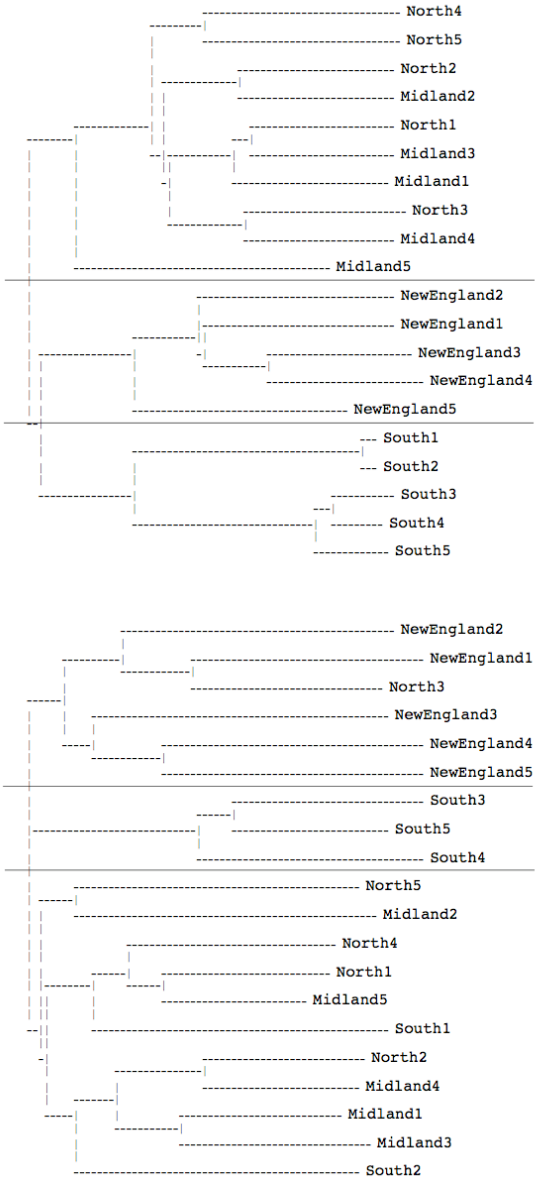
Table 1: Summary of the classification strategies for the native and non-native listeners.

	Natives	Non-Natives
Number of Groups	6.11	6.25
% Correct Pairings	43	25
% Pairwise Errors	10	17

Separate talker similarity matrices were constructed from the free classification data for the native and non-native listeners 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. These similarity matrices were submitted to the additive similarity tree algorithm ADDTREE [4]. Figure 2 shows the similarity trees for the native (top) and non-native (bottom) listeners. In these figures, perceptual distance is indicated by the sum of the lengths of the fewest number of horizontal branches required to connect any two talkers, and vertical distance is irrelevant.

The perceptual similarity structures of both the native and non-native listeners revealed three clusters of talkers. For the native listeners, the clusters perfectly differentiated the Southern (bottom), New England (middle), and North and Midland (top) talkers and the fit of the model was very good (r^2=.96). For the non-native listeners, the clusters differentiated three of the Southern talkers (middle), all of the New England and one of the Northern talkers (top), and all of the other talkers, including all of the Midland talkers and the remaining Northern and Southern talkers (bottom). The overall fit of the model was somewhat lower for the non-native (r^2=.73) than the native listeners. These results suggest that the two listener groups had similar perceptual dialect similarity structures, including three broad dialect categories (New England, South, and North and Midland), but that these categories were more clearly defined for the native than the non-native listeners.

Figure 2: Additive similarity tree solutions for the native (top) and non-native (bottom) listeners.



To explore the acoustic correlates of the clusters that emerged from the additive similarity tree analysis, a series of point-biserial correlations was conducted on the acoustic measures from [2] and the cluster affiliations shown in Figure 2. As shown in Table 2, r-fulness in the word *dark* was significantly negatively correlated with the New England cluster for both the native and the non-native listeners, s-voicing and s-duration in *greasy* were significantly correlated with the Southern cluster for both the native and non-native listeners, and s-voicing in *greasy* was significantly negatively correlated with the North and Midland

cluster for both the native and the non-native listeners. In all cases, the correlations were somewhat stronger for the non-native than the native listeners, suggesting that the native listeners may have been relying on additional acoustic correlates of regional dialect not examined in [2] to make their classification judgments. This interpretation is further supported by a closer examination of the clustering of the Southern talkers by the native and non-native listeners. The non-native listeners differentiated between the Southern talkers who said *greazy* [grizi] in the South cluster and the Southern talkers who said *greasy* [grisi] in the North and Midland cluster, whereas the native listeners grouped all of the Southern talkers together in the Southern cluster. This finding suggests that the native listeners were able to use other cues in the acoustic signal as well as signal-independent knowledge about how features cluster together across dialects to group all of the Southern talkers together, whereas the non-native listeners relied more heavily on the [grisi]-[grizi] alternation to differentiate regional dialects.

Table 2: Correlations between acoustic measures and perceptual dialect clusters. *p < .05, **p < .01

	Natives	Non-Natives
North & Midland s-voicing in *greasy*	r = -.47*	r = -.55*
New England r-fulness in *dark*	r = -.54*	r = -.63**
South s-voicing in *greasy*	r = .69**	r = .98**
South s-duration in *greasy*	r = -.48*	r = -.81**

4. DISCUSSION

Non-native listeners could use lawful variation in the acoustic speech signal to classify talkers by regional dialect of American English. They made the same number of groups of talkers, exhibited a similar perceptual similarity structure with three primary dialect clusters, and attended to the same acoustic dimensions as the native listeners. However, the non-native listeners made fewer correct talker pairings overall than the native listeners, suggesting that perceptual dialect classification is also significantly affected by linguistic experience.

The results of the correlation analysis between the clustering results and the acoustic measures suggest that the non-native listeners based their classification on a more purely sound-based set of criteria, since they had no knowledge of how the sound-based features group together in the different dialects. Thus, while dialect differences are perceptually quite salient, knowing how the features cluster together across dialects requires experience with the language and, ultimately, acquisition of a new set of indexical categories.

Differences in performance between the native and non-native listeners may also reflect different perceptual sensitivities to specific phonological contrasts. The acoustic features that were found to correlate with perception in this experiment were nearly-categorical consonant features (e.g., r-lessness vs. r-fulness, [grisi] vs. [grizi]). However, dialects of American English are typically defined with respect to gradient vowel variation. Additional research is needed to explore the role of specific linguistic variables in the perceptual similarity structure of dialect variation by native and non-native listeners.

5. ACKNOWLEDGMENTS

This work was supported by NIH F32 DC007237 and NIH R01 DC005784 to Northwestern University. The authors would like to thank Jennifer Alexander and Judy Song for their assistance with data collection.

6. REFERENCES

[1] Clopper, C.G., Levi, S.V., Pisoni, D.B. 2006. Perceptual similarity of regional varieties of American English. *J. Acoust. Soc. Am.* 119, 566-574.

[2] Clopper, C.G., Pisoni, D.B. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *J. Phonetics* 32, 111-140.

[3] Clopper, C.G., Pisoni, D.B. in press. Free classification of regional dialects of American English. *J. Phonetics*.

[4] Corter, J.E. 1982. ADDTREE/P: A Pascal program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. *Behavior Research Methods and Instrumentation* 14, 353-354.

[5] Fisher, W.M., Doddington, G.R., Goudie-Marshall, K.M. 1986. The DARPA speech recognition research database: Specification and status. In *Proceedings of the DARPA Speech Recognition Workshop,* 93-99.

[6] Imai, S., Garner, W.R. 1965. Discriminability and preference for attributes in free and constrained classification. *J. Exp. Psychol.* 69, 596-608.

[7] Stephan, C. 1997. The unknown Englishes? Testing German students' ability to identify varieties of English. In: Schneider, E. W. (ed), *Englishes Around the World*. Amsterdam: John Benjamins, 93-108.

[8] Strange, W. (ed.) 1995. *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York.