

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

**Some Acoustic Cues for Categorizing American English Regional Dialects: An
Initial Report on Dialect Variation in Production and Perception¹**

Cynthia G. Clopper

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹ This work was supported by the NIH-NIDCD R01 Research Grant DC00111 and the NIH-NIDCD T32 Training Grant DC00012 to Indiana University. I would like to thank Caitlin Dillon for her assistance in selecting the talkers for this project, Luis Hernandez for his technical assistance and support, Dr. Kenneth deJong for his assistance in selecting the measures for the first experiment, Dr. Robert Nosofsky for his assistance in conducting the Similarity Choice Model and ADDTREE analyses of the data in the second experiment, Dr. David Pisoni for his help and encouragement throughout all stages of this project, and Jimmy Harnsberger for his comments on earlier versions of this paper.

Some Acoustic Cues for Categorizing American English Regional Dialects: An Initial Report on Dialect Variation in Production and Perception

Abstract. Phonological differences between regional dialects of American English are well established in the sociolinguistics literature. The perception of these phonological differences by naïve listeners is much less well understood, however. Using an existing corpus of spoken sentences produced by talkers from a number of distinct regional dialects in the United States, an acoustic analysis was conducted in Experiment I to confirm that certain phonetic features differentiate the dialects. Results provided further evidence for predictable phonological differences between dialects. In Experiment II recordings of the sentences were played to naïve listeners who were asked to categorize each talker into one of six geographical dialect regions. Results suggested that listeners are able to reliably categorize talkers into three broad dialect clusters, but have more difficulty accurately categorizing talkers into six smaller regions. Correlations between the acoustic measures and both actual dialect affiliation of the talkers and dialect categorization of the talkers by the listeners revealed that the listeners in this study were, for the most part, able to reliably use acoustic-phonetic features of the dialects in categorizing the talkers. Taken together, the results of these experiments suggested that naïve listeners are sensitive to phonological differences between dialects and can use these differences to categorize talkers by dialect.

Introduction

Studies of regional dialects in the United States tend to focus on either phonological descriptions of specific dialects or on social aspects of attitudes towards certain dialects, such as perceived “correctness” or stereotypes related to speakers of a given dialect (e.g., Labov, Ash, & Boberg, 1997; Preston, 1986; Preston, 1989; Preston, 1993; Wolfram & Schilling-Estes, 1998). The main focus of phonological investigations of regional dialects of American English is generally the vowel system. The current shift in the vowel systems of two regions in particular has received much attention in the past decade: the Northern Cities vowel shift and the Southern vowel shift. The Northern Cities vowel shift is characterized by a clockwise rotation of the low vowels in the vowel space as shown on the left in Figure 1 and is found in such urban areas as Buffalo, Cleveland, Detroit, and Chicago. The Southern vowel shift, on the other hand, is characterized by a centralization of the tense high vowels and the lengthening of the lax high front vowels as shown on the right in Figure 1. This shift is found more prominently in rural areas of the South, as opposed to the more urban populations that exhibit the Northern Cities vowel shift. A third phenomenon involving vowels in American English that has received attention in the literature is the Low Back Merger in which /ɔ/ and /ɑ/ have merged to make homophones of such pairs as “caught” and “cot” or “Dawn” and “Don.” This merger is found in the Midland areas and much of the West, but does not appear to extend to California (Wolfram & Schilling-Estes, 1998).

Labov and his colleagues (1997) have been working on a more complete phonological description of American English, using data collected from telephone surveys of over 600 talkers across the country. The recordings from these talkers are impressionistically transcribed and acoustic measurements of F1 and F2 are taken for each of the vowels they selected to study. Based on the differences in vowel production, the preliminary Phonological Atlas of North America identifies various levels of dialect boundaries that range from a basic North-South-West split to the division of New England into Eastern New England, Western New England, and New York City.

While vowels have been the primary focus of phonological dialect descriptions, such consonantal phenomena as the post-vocalic r-lessness found in New England and some parts of the South, and the “greasy” ~ “greazy” alternation found in the South have also been noted features in discussions of phonological differences (Wolfram & Schilling-Estes, 1998).

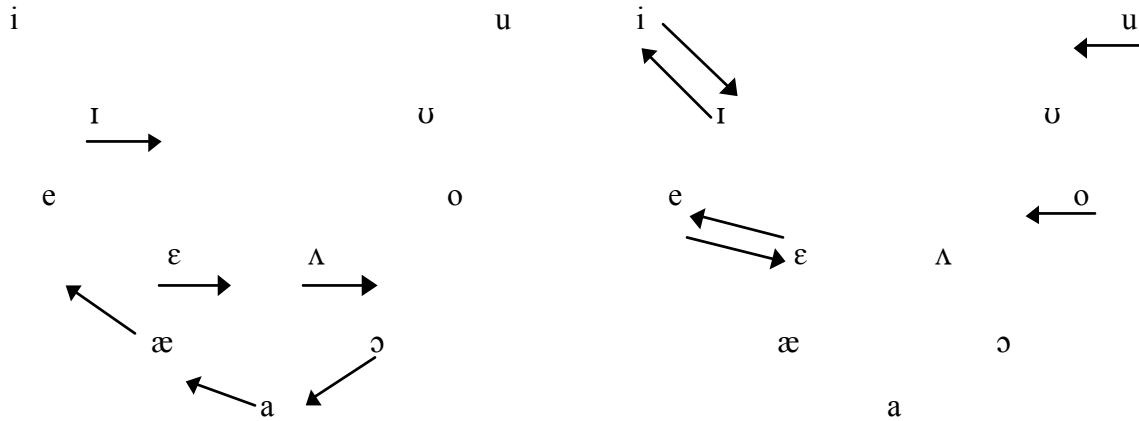


Figure 1. Northern Cities Vowel Shift (left) and Southern Vowel Shift (right). Adapted from Wolfram and Schilling-Estes (1998, pp. 138-139).

When it comes to perceptual work on dialect variation, few studies have been aimed at eliciting data from listeners based on actual speech samples. For example, Preston (1986; 1989) conducted a series of studies in which he asked undergraduates from various parts of the country to complete a number of tasks, including drawing and labeling dialect regions on a map of the United States, ranking all 50 states and a couple of key cities (New York City, Washington, D.C.) on the “correctness” or the “pleasantness” of the English spoken there, etc. Results of the map-drawing studies, conducted in Hawaii, southern Indiana, eastern Michigan, New York City, and western New York, indicated that undergraduates cannot accurately duplicate the dialect boundaries drawn by such researchers as Labov. Comparison between the composite maps of each group indicated that concepts of dialect variation are in part related to where one lives. In general, regions in close geographic proximity to any one respondent group were more finely delineated than regions farther away. It is also interesting to note that in all of the composite maps for these groups, there was at least one area on each map that was not identified as being part of any dialect region (Preston, 1986). Results of the ranking task for informants in southern Indiana indicated that “pleasantness” seems to correspond to geographic proximity to Indiana, whereas “correctness” seems to correspond more to stereotypes of where “standard” English is spoken, with California and the North and Northeast regions receiving the highest rankings (Preston, 1989).

There are two notable exceptions when it comes to the paucity of research involving behavioral responses to speech samples in regional dialect identification. The first is a recent study by Niedzielski (1999) involving listeners from Detroit who were asked to select from a set of six synthetic vowels the one that was the closest match to a vowel produced by a single female talker. One group of listeners was told that the talker was Canadian, while another group was told that the talker was from Michigan. The

results indicated that the listeners who were told that the talker was from Michigan more often selected canonical vowels as the matching vowels, while the listeners who were told that the talker was Canadian more often selected the actual matching vowels. Niedzielski concluded that a priori knowledge of a talker's dialect can affect perception of that talker's speech, particularly in terms of vowel space.

The second study, conducted by Preston (1993), considered the relationship between speech perception and dialect identification from a different perspective. Specifically, undergraduates in Michigan and Indiana were asked to listen to short speech samples taken from interviews with middle-aged males and to assign the different voices to one of nine regions, running north to south between Saginaw, MI and Dothan, AL. Results of this study revealed that the listeners were only able to make broad distinctions between North, South, and Midland. Preston noted that these perceptual boundaries did not correspond to the boundaries drawn by these same listeners in the map-drawing task discussed above. It is also interesting to note that the boundaries perceived by the Indiana residents were different from those perceived by the Michigan residents. Again, it seems that where one lives has an impact on one's perceptions of dialect variation.

While Preston's (1993) study provided some interesting insight into how listeners actually perceive dialectal differences and how those perceptions relate to geographical identification of a talker's home, no one has continued this line of research. The present experiments were designed to identify the acoustic cues that are used by listeners in identifying where a talker is from. Wolfram and Schilling-Estes claim that, "phonological patterns can be diagnostic of regional and social differences, and a person who has a good ear for dialects can often pinpoint a talker's general regional and social affiliation with considerable accuracy based solely on phonology" (1998, p. 67). However, there is little, if any, experimental evidence available to explain how listeners are able to use this knowledge of variation in phonological patterns as a diagnostic for regional identification. Even if the claim that "Southerners are more readily identified as Southerners by their /ay/ vowels than by any other single dialect feature..." (Wolfram & Schilling-Estes, 1998, p. 75) is correct, it would be useful to determine what specific phonetic features discussed in the phonological literature on dialects are actually used by naïve listeners in identifying regional dialects of American English. The goal of the present research was to investigate dialectal variation in both production and perception. Specifically, Experiment I assessed the reliability of some acoustic cues in distinguishing between talkers from different dialects. Experiment II assessed the ability of naïve listeners to use those acoustic cues in categorizing the same set of talkers by dialect region.

Experiment I: Acoustic Analysis

Methods

Talkers. Sixty-six talkers were selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Zue, Seneff, & Glass, 1990). The TIMIT corpus consists of recordings of 630 talkers reading 10 sentences each. The corpus includes 438 males and 192 females, and the talkers were each given one of eight regional labels to indicate their dialect: New England, North, North Midland, South Midland, South, West, New York City, or Army Brat. While this database was initially designed for use in speech recognition research, it has been used in a number of phonetic studies looking at the role of gender, dialect, and age in language variation (e.g. Byrd, 1992; Byrd, 1994; Keating, Blankenship, Byrd, Flemming, & Todaka, 1992; Keating, Byrd, Flemming, & Todaka, 1994). Until the present study, it has not been used in perceptual research on dialect variation.

The sixty-six talkers selected for this phonetic study were all white males who were between the ages of 20-29 at the time of recording. Eleven talkers were chosen from each of six dialect regions: New England, North, North Midland, South Midland, South, and West. The talkers were selected by the author and a second phonetically trained listener by first eliminating those talkers who did not meet the age, gender, and race requirement for each of the six dialects. Eleven talkers were then selected from each region based on repeated listening to all ten sentences spoken by each talker such that those chosen shared the most features predicted by their dialect label. Specifically, all of the New England talkers selected were r-less. The Northern talkers were selected based on their degree of /æ/ raising and /ou/ fronting. South Midland and Southern talkers selected produced monophthongal /aɪ/. Some Southern speakers also produced fronted /u/ or a merger of /ɛ/ and /ɪ/. The Western speakers who were selected all produced fronted /u/ and some also displayed the merger of /ɛ/ and /ɪ/ or a merger of /a/ and /ɔ/. Finally, the North Midland speakers selected produced none of the characteristic features of the other five dialects.

Stimulus Materials. Of the ten sentences spoken by each talker in the TIMIT database, two of the sentences were read by all of the talkers. These two “calibration sentences” were written to include specific phonemes in certain phonetic contexts in which dialect variation would be predicted (Zue et al., 1990). These two calibration sentences were used in this experiment and are shown in (1) below:

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

Each sentence for each talker was contained in a separate sound file that was segmented to include only the sentence material. For the purposes of analysis, the sound files were all leveled to 55 dB using Level16 (Tice & Carrell, 1998).

Procedure. Eleven acoustic measures were obtained from the two calibration sentences for each of the sixty-six talkers and are shown in Table 1. All of the measurements were made using Syntrillium's CoolEdit 96 program. The duration measurements were made directly from the spectrograms. Formant frequency measurements were made using the frequency analysis tool in CoolEdit 96, with a 1024 point Hamming FFT window. Frequency measurements taken at the “midpoint” were taken at the temporal midpoint of the vowel. Frequency measurements taken at the “onset” were taken at the temporal point marking the first third of the vowel. Frequency measurements taken at the “offset” were taken at the second to last glottal pulse of the vowel. All frequency measurements were taken at the peak of a glottal pulse.

In order to provide a means of normalizing frequency measures across the different talkers, the maximum F2 in the word “year” was measured for each talker. The motivation for selecting this particular measure is that the maximum F2 in the vowel /i/ in “year” should indicate the front-most edge of a given talker's vowel space. Comparing this measure to the F2 measures of other vowels can be used to determine the relative backness of those other vowels in the talker's space. Given that all of the talkers used in this experiment were male, the differences due to vocal tract size should be minimal, but taking relative backness measures instead of absolute backness measures should provide a less noisy data set.

The eleven acoustic measures were selected because we expected that they would demonstrate differences between the six dialect regions in terms of production. Four of these measures were obtained from consonants and the remaining seven from vowels. Of the seven vowel measures, three assessed vowel backness and four assessed degree of diphthongization.

New England talkers and some Southern talkers are r-less (Wolfram & Schilling-Estes, 1998). It was predicted that the F3 transition for those talkers would be smaller than for the talkers from the remaining four dialects. As a measure of r-fullness, the F3 transition in “dark” was measured by subtracting F3 at the offset of the vowel from F3 at the midpoint of the vowel.

Two alternations were predicted to distinguish the South and South Midland talkers from the other four dialect groups. An alternation between “wash” and “warsh” is found in some Southern and South Midland talkers. This epenthetic r has the effect of darkening the preceding vowel. We therefore predicted that the Southern talkers, and perhaps the South Midland talkers, should have darker vowels in “wash” than talkers from the other dialects. In order to provide some measure of the effect of this alternation on the brightness of the preceding vowel, the midpoint of F3 in “wash” was measured. There is also a “greasy” ~ “greazy” alternation that occurs in Southern and South Midland speech (Wolfram & Schilling-Estes, 1998). It was predicted that talkers from the South and South Midland would have a greater voiced proportion of the fricative in the word “greasy” and that the fricative duration would be shorter relative to the length of the entire word than for talkers from other dialect regions. This voicing alternation was measured in two ways. The first was the proportion of the fricative that was voiced. The second was the ratio of the duration of the entire fricative to the duration of the entire word.

Word	Segment	Measurement	Acoustic-Phonetic Property
dark	/a/	F3 midpoint – F3 offset	r-fullness
wash	/a/	F3 midpoint	vowel brightness
greasy	/s/	proportion of fricative that is voiced	fricative voicing
		ratio of fricative duration to word duration	fricative duration
suit	/u/	maximum F2 in “suit” – F2 midpoint	/u/ backness
don’t	/ou/	maximum F2 in “suit” – F2 midpoint	/ou/ backness
		F2 midpoint – F2 offset	/ou/ diphthongization
rag	/æ/	maximum F2 in “suit” – F2 midpoint	/æ/ backness
		F2 offset - F2 onset	/æ/ diphthongization
like	/aɪ/	F2 offset – F2 midpoint	/aɪ/ diphthongization
oily	/oɪ/	F2 offset – F2 midpoint	/oɪ/ diphthongization

Table 1. Acoustic measures selected for comparison between dialect groups

Southern talkers also produce more fronted /u/ vowels, relative to the northern dialect regions (Wolfram & Schilling-Estes, 1998). Western talkers also demonstrate a similar trend of fronted /u/ productions (Labov et al., 1997). Western and Southern talkers were therefore predicted to have fronted /u/’s and therefore have smaller relative backness values than talkers from the other regions. Northern talkers tend to produce more rounded /ou/’s than talkers from the other regions, and this should be reflected in a greater relative backness value for those talkers (Labov et al., 1997). The relative backness of the /æ/ vowel should be smaller for Northern talkers than for any of the other regions due to the upward and forward movement of /æ/ as part of the Northern Cities vowel shift (Wolfram & Schilling-

Estes, 1998). The relative backness of these three vowels was measured in the words “suit,” “don’t,” and “rag” for each talker. The midpoint of F2 in “suit” was measured and then subtracted from the maximum F2 in “year” to obtain a relative backness value of the /u/ vowel. Similarly, the midpoints of F2 in “don’t” and “rag” were measured and then subtracted from the maximum F2 in “year” to obtain relative backness values for the vowels /ou/ and /æ/.

The diphthongization measure for the /ou/ in “don’t” was also predicted to separate the Northern talkers from the others, because Northern talkers typically show less diphthongization of this vowel (Labov et al., 1997). Similarly, Southern talkers were expected to show less diphthongization of the /aɪ/ in “like” and the /oɪ/ in “oily,” given that there is a tendency for these talkers to produce monophthongal /aɪ/ and /oɪ/ (Wolfram & Schilling-Estes, 1998). There is also some evidence that the /æ/ in “rag” is becoming diphthongized in certain urban regions in the northeast (Labov et al., 1997). Based on this observation, it was predicted that greater diphthongization would be found for this vowel in the speech of New England, and possibly Northern, talkers. Measures of diphthongization were taken by subtracting the offset of F2 from the midpoint of F2 in each of the vowels. In the case of /æ/, the diphthong was measured by subtracting the offset of F2 from the onset of F2, in order to magnify any potential differences between dialect groups.

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

Results and Discussion

The acoustic analysis confirmed that there are consistent differences in speech production between the six dialects on a number of the acoustic measures considered in this analysis. The means for each of the measures are shown for each dialect group in Table 2. A series of one-way ANOVA’s was performed to determine which acoustic measures of speech production reliably distinguish between talkers of different dialects. The r-fullness measure was significant ($F(5, 60) = 3.4, p < 0.01$), as were the fricative voicing measure ($F(5, 60) = 7.2, p < 0.001$), the fricative duration measure ($F(5, 60) = 4.0, p < 0.01$), the /u/ backness measure ($F(5, 60) = 6.6, p < 0.001$), the /ou/ diphthongization measure ($F(5, 60) = 3.8, p < 0.01$), and the /æ/ backness measure ($F(5, 60) = 3.6, p < 0.01$). Means of the remaining five measures, vowel brightness, /ou/ backness, /æ/ diphthongization, /aɪ/ diphthongization, and /oɪ/ diphthongization were not significantly different.

Post-hoc Tukey tests revealed that New England differed significantly from South Midland and West on mean r-fullness ($p < 0.01$). The mean fricative voicing value for New England differed significantly from South ($p < 0.01$). The mean fricative duration value for North differed significantly from South ($p < 0.01$). The mean value of /u/ backness for New England differed significantly from South Midland, South, and West, and /u/ backness was also significantly different between North and South Midland (all $p < 0.01$). Degree of /ou/ diphthongization was significantly different for North and South. Finally, New England and North were significantly different in terms of /æ/ backness.

	New England	North	North Midland	South Midland	South	West
r-fullness (Hz)	262	409	358	462	422	451
vowel brightness (Hz)	2373	2302	2330	2133	2203	2179
fricative voicing (%)	.07	.05	.02	.27	.57	.03
fricative duration (%)	.33	.36	.36	.34	.29	.35
/u/ backness (Hz)	609	557	496	293	337	334
/ou/ backness (Hz)	1004	1105	991	1038	1012	939
/ou/ diphthong (Hz)	-71	-148	-40	22	37	-41
/æ/ backness (Hz)	601	399	440	425	494	491
/æ/ diphthong (Hz)	256	177	255	280	223	233
/aɪ/ diphthong (Hz)	452	418	402	278	331	350
/oɪ/ diphthong (Hz)	301	384	434	250	226	445

Table 2. Summary of means of acoustic measurement.

In order to determine how well a talker's dialect affiliation is associated with the acoustic properties measured in production, a series of point biserial correlations was performed. For each talker, the value on each acoustic measure (on a continuous scale) was correlated with dialect affiliation. Dialect affiliation was quantified dichotomously, such that the eleven talkers from a given dialect were given a value of "1" for that region and the remaining fifty-five talkers were given a value of "0" for that region. Results of these correlations are shown in Table 3. These correlations indicate that, as predicted, r-lessness is associated with New England talkers. New England talkers also have a greater degree of backness in /u/'s and /æ/'s, which was an unpredicted result. South Midland talkers have fronted /u/'s, which was predicted for the Southern talkers. By contrast, Southern talkers have predictably high amounts of fricative voicing in "greasy" and a predictably short fricative in the same word, but the South Midland talkers do not. Northern talkers display the predicted monophthongal /ou/. North Midland and West talkers do not show any strongly predictable measures from this analysis. Additionally, the measures of vowel brightness, /ou/ backness, and all three diphthongs did not distinguish any of the dialect groups. These correlations suggest that while many of the measures differ in their means between dialects, only a handful are truly associated with a talker's dialect affiliation. While these acoustic properties can be associated with dialect regions, they are not necessarily the only features, or the most important features, of that dialect region. The data analyzed in this experiment suggest only that some of these properties can be associated with dialect affiliation. The acoustic measures associated with dialect affiliation are, therefore, "characteristic features" of that dialect.

The results of this acoustic analysis confirm that these talkers can be reliably distinguished by dialect based on a handful of consistent acoustic differences in speech production. The following perceptual experiment was designed to investigate how well naïve listeners can use these consistent differences to categorize talkers by dialect based on short speech samples.

	New England	North	North Midland	South Midland	South	West
r-fullness	-.41**	.05	-.11	.21	.09	.18
vowel brightness	.24	.10	.16	-.25	-.10	-.15
fricative voicing	-.13	-.17	-.20	.14	.55**	-.19
fricative duration	-.08	.23	.16	-.01	-.44**	.14
/u/ backness	.38*	.26	.13	-.32*	-.22	-.23
/ou/ backness	-.03	.24	-.06	.06	-.01	-.20
/ou/ diphthong	-.11	-.39**	.00	.22	.28	.00
/æ/ backness	.41**	-.25	-.11	-.16	.06	.05
/æ/ diphthong	.07	-.24	.07	.17	-.06	-.01
/aɪ/ diphthong	.23	.13	.08	-.26	-.11	-.06
/oɪ/ diphthong	-.09	.10	.21	-.20	-.25	.23

Table 3. Correlations between talker dialect affiliation and acoustic measures. N = 66 for all correlations. Correlations with significance at $p < 0.01$ are in **bold**, * indicates $p < 0.01$, ** indicates $p < 0.001$.

Experiment II: Perceptual Categorization Task

Methods

Stimulus Materials. The same stimulus materials were used in this study as in Experiment 1 above.

Listeners. Twenty-three Indiana University undergraduates served as listeners for this study. All received partial credit for an introductory psychology course for their participation. Data from five of the listeners were removed prior to analysis: 2 were non-native speakers and 3 performed statistically at chance on the task. The eighteen remaining listeners, five males and thirteen females, were all monolingual native speakers of American English with no history of hearing or speech disorders. These eighteen listeners were divided into three listener groups based on residential history. The seven listeners who had only lived in Northern Indiana (north of, and including, Indianapolis) prior to attending school in Bloomington comprised the Northern Indiana group. The five listeners who had only lived in Southern Indiana comprised the Southern Indiana group. The remaining 6 listeners had all lived out of state for some period of time prior to attending school in Bloomington and they comprised the Out-of-State group.

Procedure. The listeners were seated at personal computers equipped with KeyTec Inc. pressure sensitive activation touch screens (KTMT1315 ProE). On the screen were the six dialect regions, represented by partial maps of the United States, including state boundaries that were labeled with the name of the dialect region. The six regions are shown in Figure 2 as they were arranged on the screen. The regions were roughly 2" x 2" in dimension and adequate space was left between the regions to minimize error in the response process. Prior to beginning the experiment, the regions were displayed on the screen and the listeners were encouraged to familiarize themselves with the regions. In the first phase

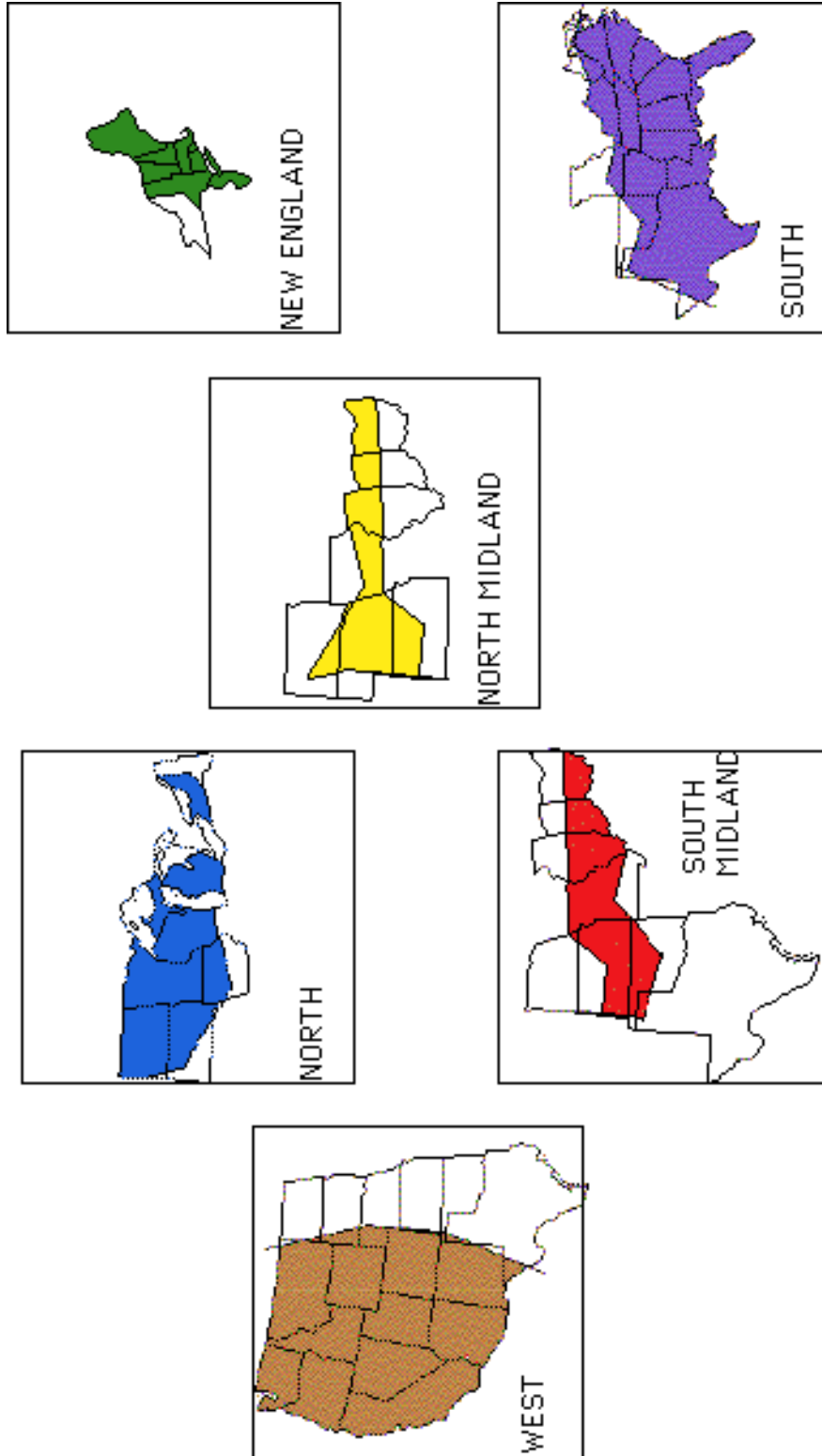


Figure 2. The six response alternatives in the categorization task. Based on Wolfram and Schilling-Estes (1998, p. 122).

of the task, the listeners responded to the first calibration sentence as spoken by each of the sixty-six talkers one time, presented in random order. On each trial, listeners heard a sentence produced by one of the sixty-six talkers, presented over headphones (Beyerdynamic DT100) at 70 dB SPL. The listeners were instructed to listen to the sentence carefully and to select the region on the screen that they thought the talker was from. The listeners made their responses by pressing directly on the screen. The listeners received no feedback about the accuracy of their responses. The second phase of the task was identical to the first, except that the listeners responded to the second calibration sentence as spoken by each of the sixty-six talkers one time, presented in random order.

Results and Discussion

Overall performance on the categorization task was quite poor. Listeners in the Out-of-State group, Northern Indiana group, and Southern Indiana group performed similarly in terms of proportion correct identification. Taken together, the three groups of listeners were only able to correctly identify where 33% of the talkers were from on the first calibration sentence and where 28% of the talkers were from on the second calibration sentence. While overall performance was low, it was statistically above chance for both sentences. The proportions of correct identifications for talkers from each dialect region for each sentence are shown in Table 4, collapsed across all three listener groups. A t-test indicated that the performance for the two sentences was not significantly different ($t(34) = 3.21, p = 0.55$).

	First Sentence	Second Sentence
New England	61	34
North	23	26
North Midland	25	27
South Midland	34	27
South	35	34
West	23	20
Mean	33	28

Table 4. Percent correct categorization of dialect affiliation of the talkers for each sentence, collapsed across the three listener groups (chance = 17%).

An inspection of the confusion matrices of responses suggested that the listeners' inability to correctly identify a majority of the talkers was not due to random responses, but was more likely due to a consistent pattern of confusions. In order to determine the structure of this pattern of errors, the 6 x 6 confusion matrices for each of the two calibration sentences for each listener group, and collapsed across all three listener groups, were submitted to the Similarity Choice Model (Nosofsky, 1985) to determine similarity and bias parameters between the dialect regions. The similarity parameters indicated the degree of similarity between each of the dialects, based on the confusion data. The bias parameters indicated the responses biases of the listeners. The bias parameters that resulted from the Similarity Choice Model analysis suggested that the listeners were not biased to respond with one alternative more or less often than any of the other response alternatives. The similarity parameters were submitted to an additive clustering scheme, ADDTREE, to determine one measure of the perceptual distances between the dialects (Corter, 1995). An additive clustering scheme was selected because the initial examination of the confusion matrices indicated that there was high reciprocity between the six regions. For example, South was most often confused with South Midland and vice versa. Other spatial analyses, such as multi-dimensional scaling, were inappropriate for this data given the small number of data points in the matrix.

The perceptual distances for the listener groups were highly correlated with each other and with the distances for all of the listener groups combined, demonstrating no significant differences between the three listener groups. All further analyses considered the data collapsed across all of the listeners. The resulting trees from the ADDTREE analysis collapsed across listener groups are shown in Figure 3. For the first calibration sentence, it is clear that listeners grouped the talkers into three main clusters: New England, South and South Midland (hereafter, South Cluster), and North, North Midland, and West (hereafter, Other Cluster). The solution for the second calibration sentence also appears to have three broad clusters: New England and North (hereafter, North Cluster), the South and South Midland (hereafter, South Cluster), and the North Midland and West (hereafter, West Cluster).

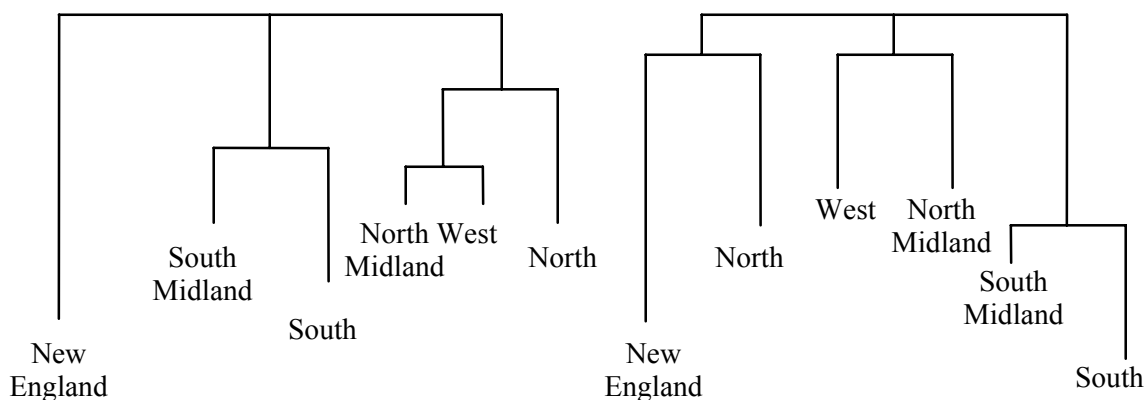


Figure 3. Clustering solution for the first (left) and second (right) calibration sentence, based on listeners' confusion matrices.

When the proportion correct categorization scores for all the listeners are collapsed into the three broad clusters for each of the two calibration sentences, performance increases dramatically, as expected. Correct categorization of talkers into New England, South Cluster, or Other Cluster for the first calibration sentence was 67%. Correct categorization of talkers into North Cluster, South Cluster, or West Cluster for the second calibration sentence was 53%. These results suggest that listeners are able to reliably categorize talkers into three broad dialect groups, rather than the six used in the TIMIT corpus. The different clustering results from the two sentences suggest that these three categories might be fluid, depending on which phonetic cues are available for identifying a talker. Recall that the first sentence contained the word “dark” and that r-lessness was a characteristic feature of the New England talkers. If listeners were able to use r-fullness as a cue in identifying talkers, it is not surprising that New England was in a cluster by itself for the first sentence when that cue was available, but that it grouped with another region when that cue was not available, as in the second sentence.

In order to determine which phonetic cues the listeners were using to categorize the talkers, a series of correlations was performed. For each talker, the value on each acoustic measure was correlated with the percent categorization of that talker into a given dialect region over all listeners. Results of these Pearson correlations are shown in Table 5. They suggest that listeners may use some of these cues in order to categorize talkers by dialect. For example, it seems that listeners can use r-lessness to identify talkers from New England, vowel darkness to identify talkers from the South Midland, fricative voicing to identify talkers from the South and South Midland, /u/ frontness to identify talkers from the South

Midland, /ou/ diphthongization to identify talkers from the South, /ou/ monophthongization to identify talkers from the North, /aɪ/ diphthongization to identify talkers from the North Midland, /oɪ/ diphthongization to identify talkers from the North Midland and the West, and /aɪ/ and /oɪ/ monophthongization to identify talkers from the South.

	New England	North	North Midland	South Midland	South	West
r-fullness	-.40**	-.06	.07	.30	.24	.02
vowel brightness	.28	-.01	.04	-.52**	-.14	.16
fricative voicing	-.16	-.28	-.27	.33*	.42**	-.13
fricative duration	.02	.12	.31	-.22	-.29	.19
/u/ backness	.31	.25	-.09	-.44**	-.31	.19
/ou/ backness	.14	.13	-.04	-.15	.02	-.17
/ou/ diphthong	-.29	-.43**	-.06	.20	.39**	-.03
/æ/ backness	.22	-.01	.14	-.19	-.15	.01
/æ/ diphthong	-.12	.01	-.10	-.07	.21	-.02
/aɪ/ diphthong	.00	.20	.37*	-.14	-.33*	.11
/oɪ/ diphthong	-.15	.21	.57**	-.22	-.45**	.45**

Table 5. Correlations between acoustic measures and dialect categorization. $N = 66$ for all correlations. Correlations with significance at $p < 0.01$ are in **bold**, * indicates $p < 0.01$, ** indicates $p < 0.001$.

The results of the two experiments taken together demonstrate that some acoustic measures are associated with a talker's dialect affiliation and that some acoustic cues are associated with how listeners categorize a given talker. In order to determine whether or not listeners use the characteristic acoustic features of the dialects in their categorization of the talkers, the correlations from the acoustic analysis have been plotted with the correlations from the perceptual experiment for each dialect region. These plots are shown in Figure 4. Plotted on the x-axis are the squared correlation coefficients from Experiment I, which reveal the proportion of variance (r^2) in the acoustic measures accounted for by the actual dialect affiliation of the talkers. Plotted on the y-axis are the squared correlation coefficients from Experiment II, which reveal the proportion of variance (r^2) in the dialect categorization of the talkers accounted for by the acoustic measures. The acoustic measures from both calibration sentences have been plotted together in these figures. If listeners used the acoustic cues optimally, the points would form a line with a slope = 1. Any points falling above the line $x = y$ represent those acoustic cues which are not characteristic features of the dialect, but which the listeners used in their categorization of the talkers. For example, in Figure 4c, the point representing degree of /oɪ/ diphthongization falls above the line $x = y$. This indicates that despite the fact that /oɪ/ diphthongization is not a characteristic feature of the North Midland dialect, listeners used this feature to discriminate North Midland talkers from other talkers. Conversely, any points falling below the line $x = y$ represent those acoustic cues which are characteristic features of the dialect, but which listeners did not use in their categorization of the talkers. For example, in Figure 4a, the point representing /æ/ backness falls below the line $x = y$. This indicates that despite the fact that /æ/ backness is a characteristic feature of New England, the listeners did not use this feature to discriminate New England talkers from other talkers.

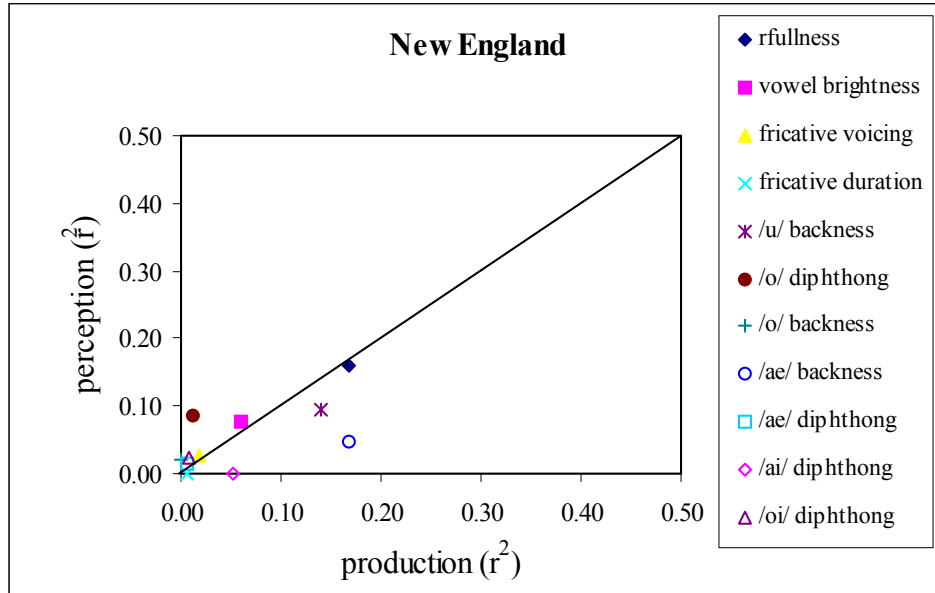


Figure 4a. Presence of features in production for both sentences v. perception by listeners in categorization for New England.

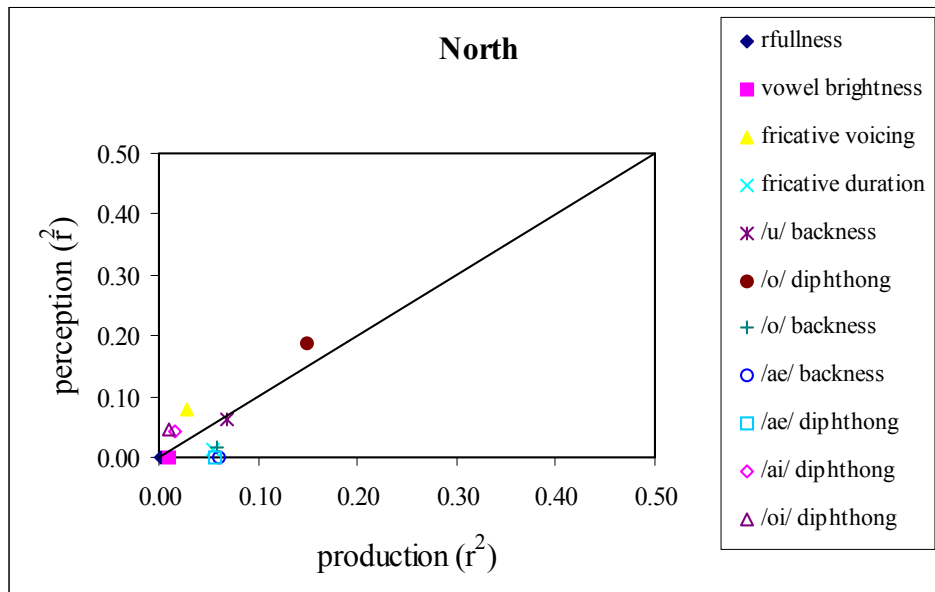


Figure 4b. Presence of features in production for both sentences v. perception by listeners in categorization for North.

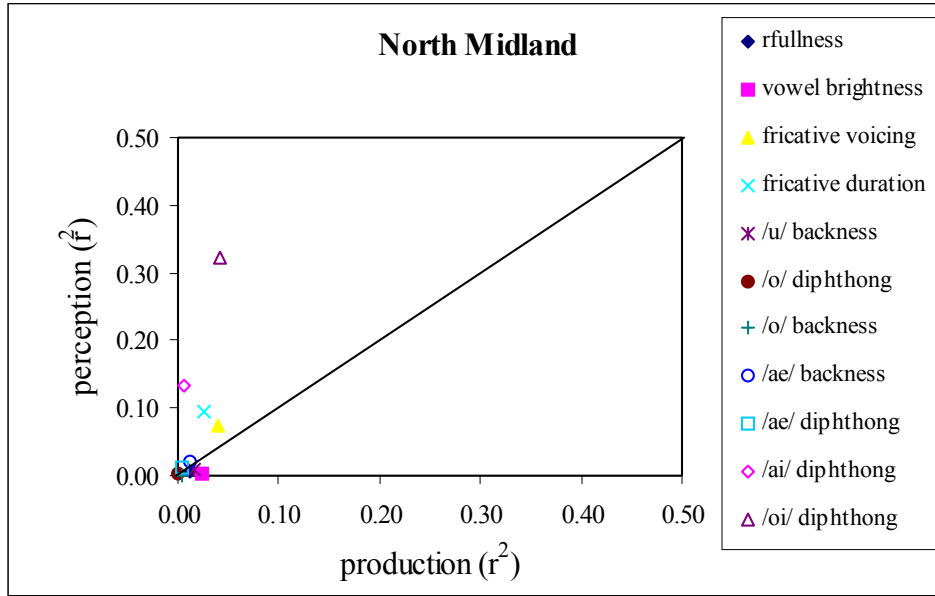


Figure 4c. Presence of features in production for both sentences v. perception by listeners in categorization for North Midland.

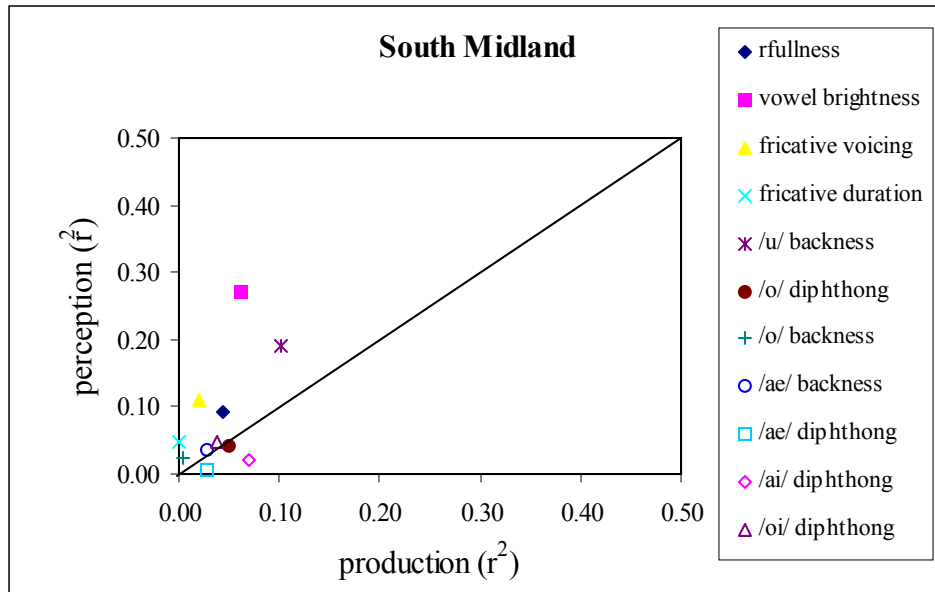


Figure 4d. Presence of features in production for both sentences v. perception by listeners in categorization for South Midland.

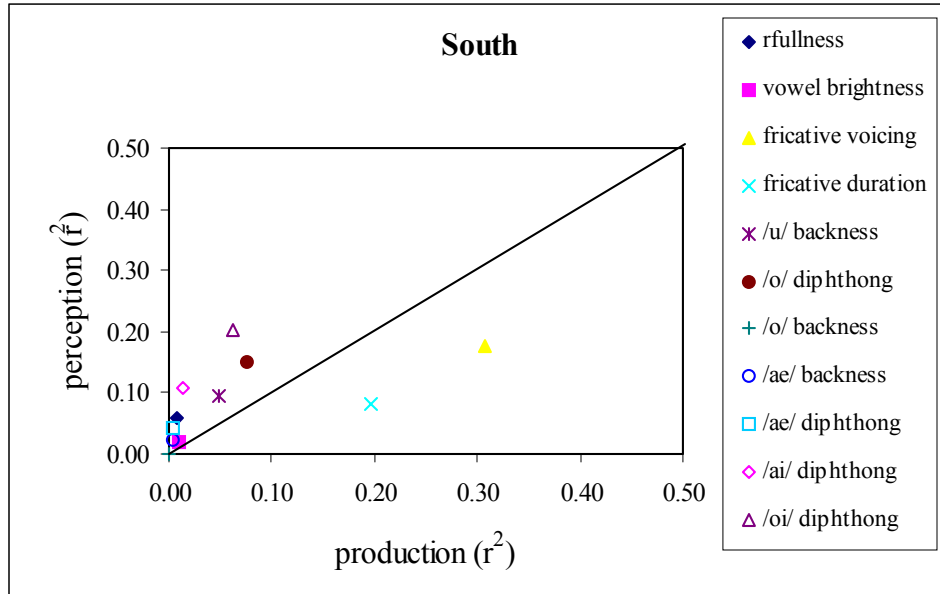


Figure 4e. Presence of features in production for both sentences v. perception by listeners in categorization for South.

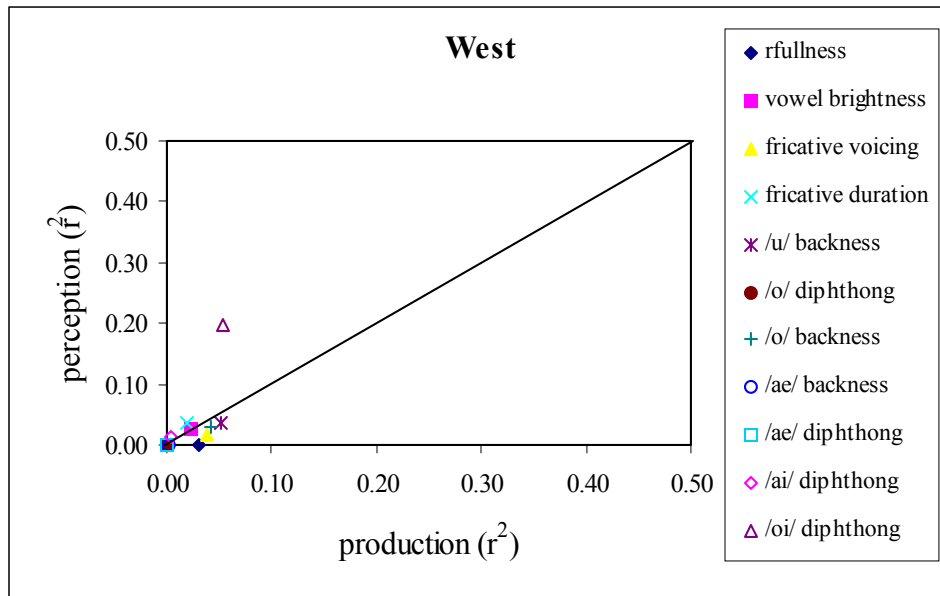


Figure 4f. Presence of features in production for both sentences v. perception by listeners in categorization for West.

These plots show several things of interest with respect to the relationship between production and perception. The first is that for all six of the dialect regions, there is a cluster of cues close to the origin. These cues are neither useful in predicting dialect affiliation nor are they used by listeners to categorize the talkers. The second notable point is that for New England, North, and South, the points not

at the origin tend to fall close to the $x = y$ line. For the North Midland, South Midland, and West, however, the points not clustered at the origin tend to fall lower on the production scale than the perception scale, indicating that the listeners were using non-characteristic features of those regions in assigning talkers to those regions. These two observations taken together indicate that listeners are much more capable of identifying and using the appropriate acoustic cues for New England, North, and South, than they are for North Midland, South Midland, and West.

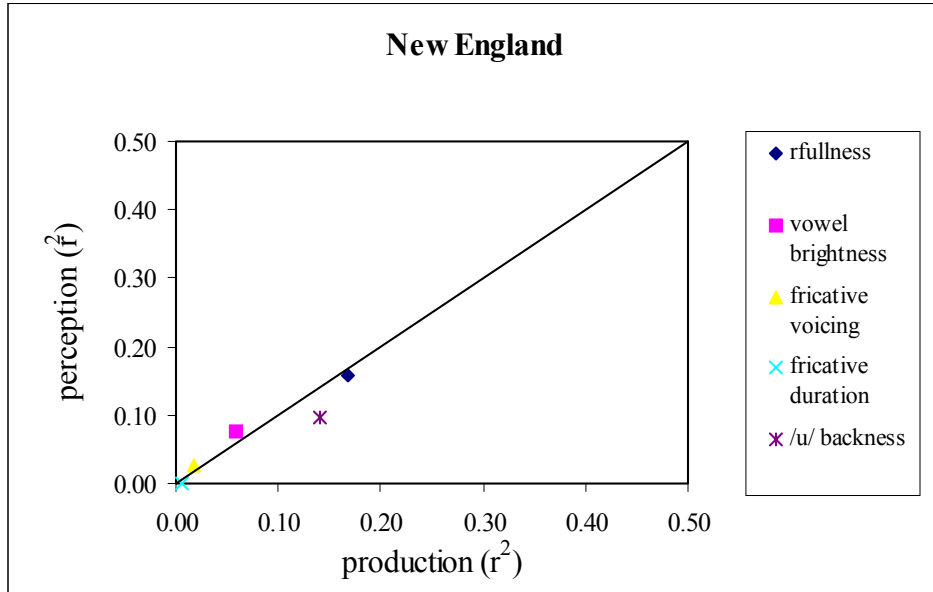


Figure 5a. Presence of features in production for the first sentence v. perception by listeners in categorization for New England.

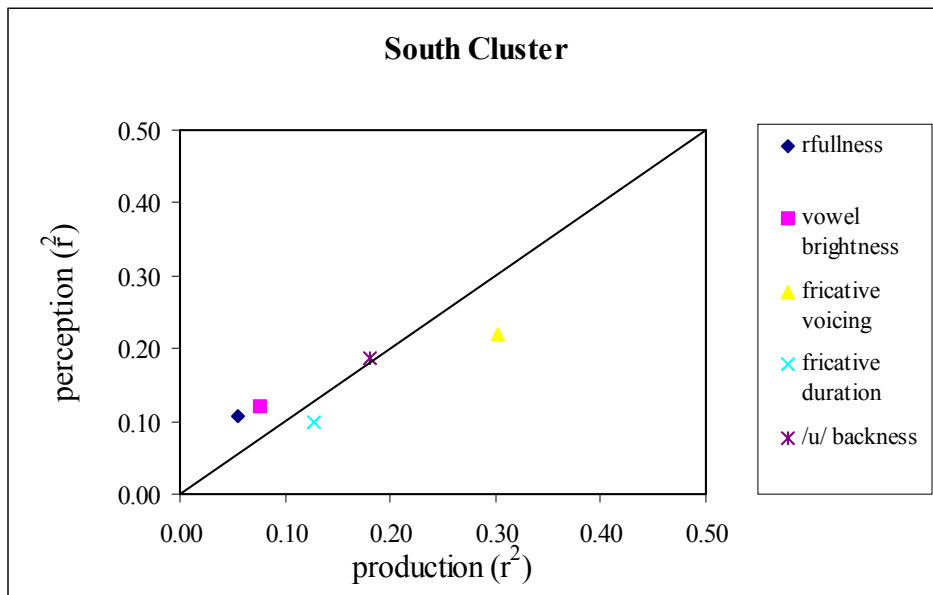


Figure 5b. Presence of features in production for the first sentence v. perception by listeners in categorization for South Cluster (South and South Midland).

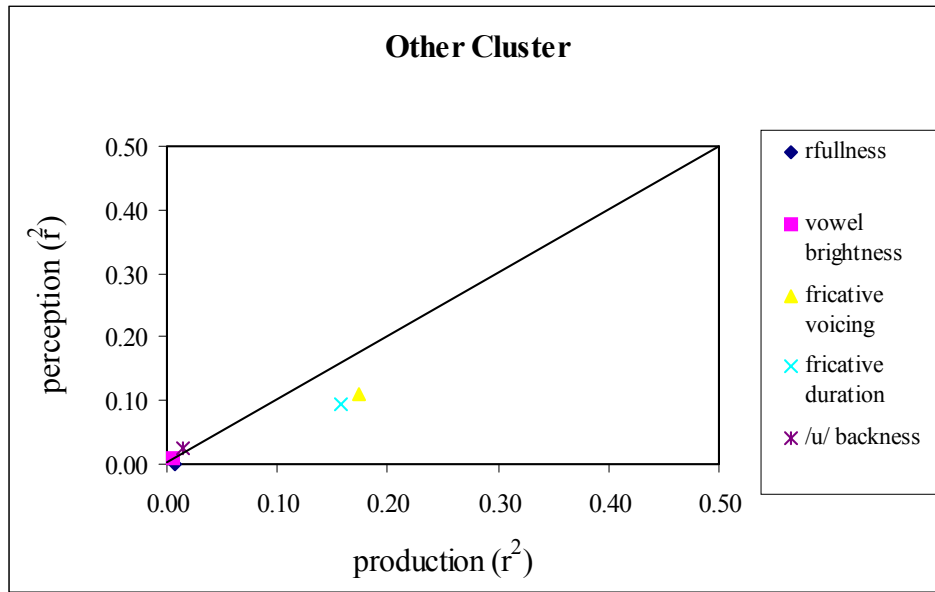


Figure 5c. Presence of features in production for the first sentence v. perception by listeners in categorization for Other Cluster (North, North Midland, West).

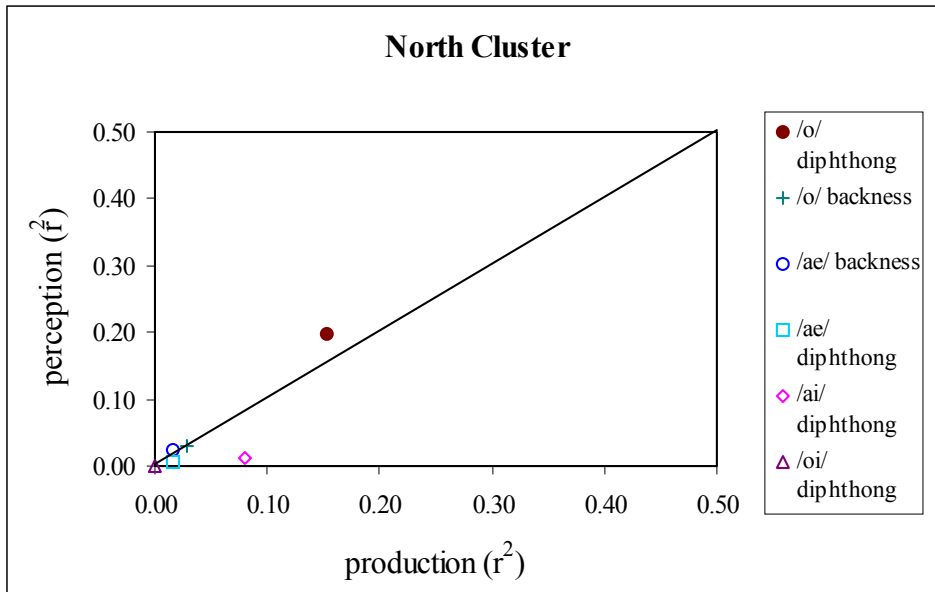


Figure 6a. Presence of features in production for the second sentence v. perception by listeners in categorization for North Cluster (New England and North).

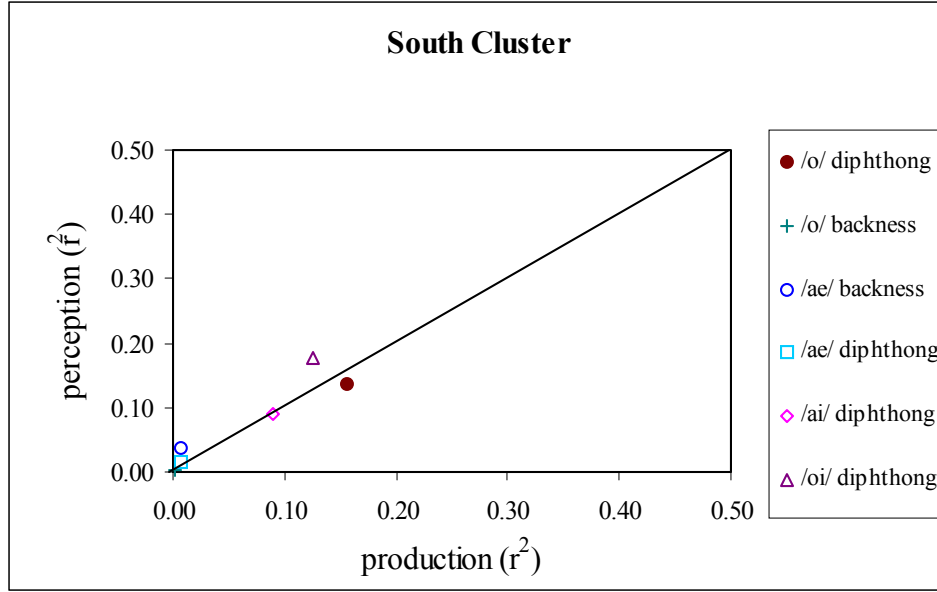


Figure 6b. Presence of features in production for the second sentence v. perception by listeners in categorization for South Cluster (South Midland and South).

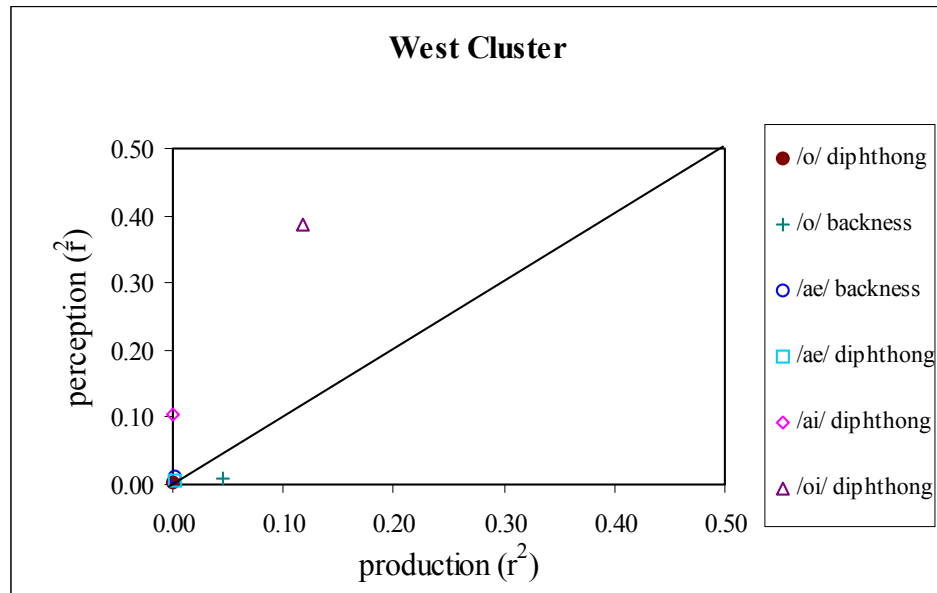


Figure 6c. Presence of features in production for the second sentence v. perception by listeners in categorization for West Cluster (North Midland and West).

One possible explanation for these differences is that the regions in the latter group are less familiar to the listeners as distinct “dialect regions.” The results of the clustering analysis above indicated that the listeners used fewer than six dialect categories reliably in this task. Therefore, a set of point biserial correlations between cluster affiliation and acoustic measures and a set of Pearson correlations between acoustic measures and percent cluster categorization were performed in the same manner as above, using data from all sixty-six talkers. The results of the point biserial correlations revealed the characteristic features of the dialect clusters and the results of the Pearson correlations revealed which cues the naïve listeners were using in categorizing the talkers by cluster. The squared correlation coefficients were then plotted against each other to provide an index of how well listeners used the acoustic cues that are good predictors of cluster affiliation. In these plots, the acoustic cues from the two sentences have been plotted separately because the clustering solutions differed for the two sentences. The plots for the first calibration sentence are shown in Figure 5 and the plots for the second calibration sentence are shown in Figure 6. This series of plots shows that listeners are relatively good at using the appropriate cues for all but the West Cluster. For the West, listeners tend to use acoustic cues that are not characteristic features of that cluster. However, none of the acoustic measures obtained in this study were highly correlated with the North Midland or the West, so it is perhaps arguable that the reason talkers from this cluster are difficult to categorize is because the cluster does not have characteristic features to distinguish it from the other regions. That is, there are no “good” cues for the listeners to use. While listeners are able to use r-lessness to identify talkers from New England and /oi/ diphthongization to identify talkers from the South, none of the acoustic properties examined in the acoustic analysis were highly associated with the talkers from the West Cluster. Therefore, the listeners may not have had any acoustic cues available to them in making their categorization judgements of the North Midland and West talkers.

General Discussion

As predicted, the acoustic analyses performed in the first experiment confirmed that, as a group, the talkers selected from each dialect reliably produce phonological differences that can be measured acoustically. Specifically, for this set of talkers, r-lessness, /u/ backness, and /æ/ backness are characteristic features of the eleven New England talkers. /ou/ monophthongization is a characteristic feature of the eleven Northern talkers. /u/ frontness is a characteristic feature of the eleven South Midland talkers. Finally, fricative voicing and duration are characteristic features of the eleven Southern talkers. None of the acoustic measures selected for this analysis were characteristic features of either the eleven North Midland talkers or the eleven West talkers.

Some of the acoustic measures that were expected to reveal differences between the dialect groups were not predictably different between the dialects. Specifically, the vowel brightness in “wash” was expected to distinguish the South and South Midland from the other dialects. However, the correlation between South Midland dialect affiliation and this measure was weak ($r = -0.25$). This measure based on F3 values is problematic, however, because it was not normalized across speakers for vocal tract size, unlike the measures involving F2 that were normalized against the F2 of “year” to account for talker differences. This measure is also potentially problematic because the vowel itself can take on a different quality in different dialects. Additionally, the measures for the diphthongs /aɪ/ and /oi/ were also predicted to distinguish the South and South Midland talkers from the others. The correlation between South Midland affiliation and the measure of /aɪ/ diphthongization was weak ($r = -0.26$) as was the correlation between South affiliation and the measure of /oi/ diphthongization ($r = -0.25$). The measure of degree of diphthongization of /aɪ/ is potentially problematic in this analysis because it was taken from the word “like.” A following velar context generally results in an upward offglide of the

preceding vowel (Ladefoged, 1993). This upward offglide may have concealed the expected monophthongization of /aɪ/ in the South and South Midland talkers. These weak associations suggest that while some of the predictions based on the current sociolinguistic literature were not entirely confirmed, there is still a relationship between some phonetic features and dialect affiliation.

Another possible explanation for the lack of correlation between some of the acoustic measures and dialect affiliation is that some of the talkers selected for this study were not good representatives of their dialect region. The standard deviations of the means shown in Table 2 reveal that there was a lot of variation between the talkers within any given dialect group. It may be the case that certain talkers in a given dialect are better representatives of their region than others. That is, some talkers may more reliably produce the phonetic features that distinguish their dialect from others and some talkers may be more easily categorized by listeners than others. Additionally, there may be some striking individual differences between the listeners that can account for some of the data presented here. Analyses of the individual talkers and the individual listeners have not been completed, but may provide some insight into why some predicted correlations did not emerge. Finally, it is possible that the regions used to define the talkers in this study are not the most accurate categorization of these talkers. For example, some recent research suggests that the Midland areas should be considered as one single region. There is also some controversy about the vast geographical area contained within the Western region (Labov et al., 1997).

The results of the categorization task in the second experiment support the findings of Preston (1993) that indicate that naïve listeners are only able to categorize talkers based on dialect into broad categories. Specifically, the listeners in this experiment were able to reliably categorize the sixty-six talkers into three broad dialect categories: North, South, and West. The placement of Northern talkers into one of these three clusters appeared to be based on the availability of r-fullness as an acoustic cue. In the first calibration sentence, the r-fullness cue was available, and the listeners used this to identify talkers from New England and placed Northern talkers in the West Cluster. In the second calibration sentence, the r-fullness cue was not available to identify New England talkers and the listeners placed Northern talkers in the North Cluster.

The listeners also demonstrated reliable use of a number of the acoustic cues in categorizing the talkers. Specifically, r-less talkers were categorized as New Englanders. Talkers with a highly diphthongal /ou/ were categorized as Northerners. Talkers with a highly diphthongal /aɪ/ and /oɪ/ were categorized as North Midlanders. Talkers with a dark vowel in “wash,” a voiced fricative in “greasy,” and a fronted /u/ were categorized as South Midlanders. Talkers with a voiced fricative in “greasy,” a highly diphthongal /ou/, and a highly monophthongal /aɪ/ and /oɪ/ were categorized as Southerners. Finally, talkers with a highly diphthongal /oɪ/ were categorized as Westerners.

Despite the consistent use of some of the acoustic cues, the listeners were not always using the most optimal cues in their decisions. That is, the most characteristic features of each dialect region, as revealed by the point biserial correlations in the first experiment, were not always the acoustic properties used by the listeners. For example, /æ/ backness was a fairly good cue characterizing New England, but the listeners did not use it. Fricative duration and voicing were also relatively good cues characterizing the South that the listeners did not use optimally. Conversely, degree of /oɪ/ diphthongization was not a good characteristic feature of North Midland, South, or West talkers, but the listeners relied heavily on this measure as an indicator of dialect region in all three cases. Similarly, vowel brightness was not a particularly good characteristic feature of South Midland talkers, but the listeners relied heavily on this cue as well.

Overall, the comparison between the two sets of correlations based on dialect regions in Figure 4 suggests that listeners used the characteristic features of New England, North, and South more optimally than those of the Midland regions and the West. The results of the clustering analysis suggested that listeners can better distinguish between three broad dialect clusters than between the six smaller regions. It is therefore reasonable to consider how well the listeners used the characteristic features of the clusters in their categorization of the talkers. The comparison between the two sets of correlations based on dialect clusters in Figures 5 and 6 suggest that listeners were in fact using the characteristic features of all of the clusters, except the West Cluster. Recall that the West Cluster is composed of the North Midland and the West regions. The results of the acoustic analyses revealed that there are no characteristic features for either of these regions in the set of phonetic features considered here. Therefore, it is not at all surprising that the listeners were relying on a feature that is not characteristic of the cluster in categorizing the talkers, because there is no reliable feature in the talkers' productions to rely on. Regardless of whether or not the listeners used the characteristic features of the dialect regions optimally in the categorization of the talkers, however, it is clear that naïve listeners are sensitive to a number of phonological differences between dialects and that extensive training is not required before listeners can use these differences to accurately identify where talkers are from, at least in terms of broad dialect clusters.

In addition to continuing to analyze the possible individual talker and listener differences in this data, this line of research can be extended in various ways. Specifically, the relatively poor performance by the listeners in the categorization task raises several issues regarding possible manipulations of the task, such as training the listeners on representative speakers of each dialect and having them generalize to new talkers or providing the listeners with a smaller set of response alternatives. Additionally, further analyses can be conducted to determine the perceptual similarities between the dialect regions and between the talkers in each region.

Conclusions

The results of the first experiment using acoustic measurement techniques provide further evidence that phonological differences do exist between regional dialects of American English and that differences in speech production can be predicted to some extent by the dialect affiliation of the talkers. The results of the second experiment provide perceptual evidence that supports Preston's (1993) findings that naïve listeners do not necessarily categorize talkers accurately by dialect region, but that they are able to make reliable distinctions between some dialect groups on a broader scale. In particular, the naïve listeners were able to reliably identify talkers from the South, the North, and New England, but they had a harder time identifying talkers from the Midland areas and the West. The results of these two experiments together suggest that listeners are aware of important phonological differences between dialects and can use their detailed knowledge to categorize talkers by dialect region, without any specific training or feedback.

References

- Byrd, D. (1992). Sex, dialects, and reduction. *ICSLP 92 Proceedings*, 827-830.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39-54.
- Cortier, J.E. (1995). ADDTREE/P Program for Fitting Additive Trees.
- Keating, P., Blankenship, B., Byrd, D., Flemming, E. & Todaka, Y. (1992). Phonetic analyses of the TIMIT corpus of American English. *ICSLP 92 Proceedings*, 823-826.
- Keating, P.A., Byrd, D., Flemming, E., & Todaka, Y. (1994). Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication*, 14, 131-142.

- Labov, W., Ash, S., & Boberg, C. (1997). A National Map of the Regional Dialects of American English. Retrieved June 26, 2000 from the World Wide Web: http://www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html.
- Ladefoged, P. (1993). *A Course in Phonetics*. Fort Worth, TX: Harcourt Brace.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*, 62-85.
- Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice-model analysis. *Perception and Psychophysics, 38*, 415-432.
- Preston, D. (1986). Five visions of America. *Language and Society, 15*, 221-240.
- Preston, D. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics*. Providence, RI: Foris.
- Preston, D. (1993). Folk dialectology. In Preston, D. (ed.) *American Dialect Research*. Philadelphia: John Benjamins, pp. 333-378.
- Tice, R. & Carrell, T. (1998). Level16 v.2.0.3. University of Nebraska.
- Wolfram, W. & Schilling-Estes, N. (1998). *American English*. Malden, MA: Blackwell.
- Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication, 9*, 351-356.

This page left blank intentionally.