

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**

Progress Report No. 26 (2003-2004)

*Indiana University*

**Some New Experiments on Perceptual Categorization of Dialect Variation in  
American English: Acoustic Analysis and Linguistic Experience<sup>1</sup>**

**Cynthia G. Clopper and David B. Pisoni**

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

---

<sup>1</sup> This work was supported by the NIH-NIDCD R01 Research Grant DC00111 and the NIH-NIDCD T32 Training Grant DC00012 to Indiana University. We would like to acknowledge the valuable advice that we received on various aspects of this project from Kenneth deJong, Caitlin Dillon, Luis Hernandez, and Robert Nosofsky, as well as the assistance of Jeffrey Reynolds and Adam Tierney with data collection.

## Some New Experiments on Perceptual Categorization of Dialect Variation in American English: Acoustic Analysis and Linguistic Experience

**Abstract.** Traditional methods of research on perceptual dialectology have been limited to tasks that ask participants to draw and label dialect regions on maps or to make attitude judgments about samples of certain linguistic varieties. Only a small handful of perceptual experiments have directly looked at how listeners identify and categorize where talkers are from based on actual speech samples. Our first experiment investigated how well naïve listeners could categorize talkers based on regional dialect of American English using sentence-length utterances. Although listeners performed poorly on this task, a more detailed analysis of their error patterns revealed systematic perceptual confusions. The results suggest that listeners have knowledge of three broad dialect categories: New England, South, and North/West. Since listeners were able to perform the perceptual categorization task at levels reliably above chance, a second experiment was carried out to measure the acoustic-phonetic properties that listeners used to make their categorization judgments. Results of this acoustic analysis study revealed four robust acoustic-phonetic properties that were good predictors of where the talkers were actually from: New England r-lessness and /æ/ backing, North /ou/ offglide centralization, and South Midland /u/ fronting. These properties also appeared to be used by the listeners in making their perceptual judgments. A third experiment explored the effects of residential history on dialect categorization performance using two different groups of listeners with different linguistic experiences. Listeners who had lived in at least three different states comprised the “Army Brat” group; listeners who had lived only in Indiana comprised the “Homebodies” group. The results revealed that the Army Brats performed significantly better than the Homebodies on the perceptual dialect categorization task. This finding suggests that greater exposure to linguistic variation in development leads to better performance on the categorization task. A final experiment investigated the effects of short-term perceptual learning on categorization performance. Using the same talkers and response alternatives as in the first experiment, one group of listeners was trained to categorize one talker from each region and a second group was trained to categorize three talkers from each region. After training, both groups of listeners were then asked to categorize new talkers using the same six dialect regions. Results showed that listeners who were exposed to a range of variability after training with three talkers from each dialect region were better able to generalize to new talkers than the listeners who were trained to identify only a single talker from each region. Taken together, these four experiments provide new evidence for the important role of acoustic-phonetic variation and variability in speech perception and spoken language processing. The implications of these findings for speech perception and spoken language processing are discussed.

### Introduction

Over a decade ago, Klatt (1989) identified five sources of variability in spoken language: ambient conditions, word environment, segment realization, within-speaker variability, and cross-speaker variability. Ambient conditions include such non-speech phenomena as background noise, room reverberation, and the properties of any recording or telephonic equipment involved. During World War II, Miller (1946) and his colleagues developed protocols for testing the effects of noise introduced through communication equipment on speech intelligibility and spoken word recognition. Models of speech perception, however, typically assume ideal listening conditions or a dual-route auditory system in which

all ambient noise is filtered out of the speech stream and processed by a separate, general auditory mechanism (Mattingly & Liberman, 1990).

Utterance-specific variability includes word environment and segment realization variability which are both due to the physical properties of the human speech production system. Word environment effects include cross-word coarticulation and durational changes due to stress, focus, and reduction processes in continuous speech. Stress, focus, and reduction also effect segment realization, as do segmental coarticulation and variable production rules (such as releasing a word-final stop). Normalization for these kinds of variability has been one of the major theoretical problems in speech perception research. Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) claimed to have found the invariant cue to consonant place of articulation in the second formant transition in CV syllables. Models of speech perception such as Klatt's (1979) Lexical Access from Spectra (LAFS) model are based on the fundamental assumption that at some level acoustic invariance can be found across utterances.

Finally, within-speaker and cross-speaker sources of variability are related to "indexical" properties of the talker: gender, age, physical size, dialect, social status, emotional state, register, and speaking rate. Emotional state, register, and speaking rate are all sources of within-speaker variability, whereas properties of the talkers such as their gender, age, physical size, dialect, and social status are considered cross-speaker sources of variability. Few researchers working in the mainstream of speech research have focused their attention on the role that these talker-specific sources of variability play in speech perception and spoken language processing and no one has attempted to model how these different sources of variation might be encoded by human listeners.

One exception to this general trend is the research from our laboratory over the last decade. Pisoni and his colleagues have examined the role of talker variability in speech perception and spoken word recognition and have discovered that indexical characteristics of a talker are actually perceived and encoded by listeners along with the meaningful content of the linguistic signal. In one series of studies, Mullennix, Pisoni, and Martin (1989) showed that cross-talker variability affected word recognition performance in noise. Listeners performed more poorly when the talker changed from trial to trial than when the talker remained constant across all trials. In addition, the results of a speeded classification task suggested that indexical properties of the talker are inseparable from the linguistic content of the utterance (Mullennix & Pisoni, 1990).

Perceptual learning studies conducted by Pisoni and his colleagues have also shown that the role of variability in spoken language stimuli in training leads to better generalization performance. For example, in a study on the perceptual learning of novel voices, Nygaard, Sommers, and Pisoni (1994) found that training listeners to identify talkers led to better performance on word recognition in noise when the words were spoken by familiar talkers they had learned to identify in the first part of the experiment than when the words were spoken by unfamiliar talkers. In another study, Logan, Lively, and Pisoni (1991) showed that training Japanese listeners to identify English /r/ and /l/ using highly variable natural stimuli led to better generalization performance to novel talkers and novel utterances than training on less variable materials. These studies provide evidence for the role of variability in speech perception and reveal that perceptual learning can be enhanced through the use of training materials that contain variability and variation.

Variability has been observed in spoken language for many years and is a natural consequence of the speech production process and the physical mechanisms used to control speech articulation. More than 50 years ago, Peterson and Barney (1952) published the results of their well-known pioneering acoustic vowel space study which revealed large amounts of variation both within and across talkers. They plotted the first and second formant values of ten vowels spoken by 76 talkers (including men,

women, and children) on an F1 by F2 plane on which they superimposed one ellipse for each of the ten vowels. The ellipses were drawn to include roughly 90% of the tokens for a given vowel. Their results produced a set of 10 relatively large and overlapping ellipses that revealed the large amount of variation in formant frequencies for vowels of the same quality. Even when only those tokens that were correctly identified by listeners 100% of the time were included in the figure, there was still a great deal of overlap between different vowels in the acoustic space. The observed pattern reflects the finding that vowels of different phonemic qualities are often found in exactly the same part of the F1 x F2 vowel space.

Despite this overwhelming evidence for variation and variability in vowel production, however, researchers working on human speech perception and speech synthesis typically focused on the mean formant values reported by Peterson and Barney (1952). For example, Hillenbrand, Getty, Clark, and Wheeler (1995) carried out an extensive study that was designed to replicate Peterson and Barney's vowel spaces using a larger number of talkers and a larger corpus of vowels. Hillenbrand et al. were interested in replicating and extending the mean formant values found in the earlier study. What they found instead was a systematic shift in the low front vowels that reflects the Northern Cities vowel shift in Michigan and other northern states where the new recordings were made. Peterson and Barney had recorded their talkers forty years earlier and had not controlled for regional variety of American English or even for native language. Hagiwara (1997) noted these differences in mean formant frequency between the two studies and pointed out the obvious role that four decades and geographic location played in the differing results. He then replicated the Peterson and Barney study again with talkers from Southern California. He found a shift in the back vowels that reflects the back vowel fronting found in the southern United States and in some parts of California. Hagiwara argued that speech researchers should work to record and measure vowel spaces of talkers in different parts of the country in order to determine the extent of vowel production variability.

Sociolinguists have been documenting precisely this kind of phonological variation in speech for more than thirty years, since Labov revolutionized the field with quantitative methods for collecting spoken language data on variable productions in his famous New York City department store study (Labov, 1972). More recently, Labov, Ash, and Boberg (in press) have collected recordings of over 700 talkers from around the United States and carried out acoustic measurements of the vowels. These measurements have allowed them to determine current dialect boundaries in the United States and to describe current changes in progress such as the Northern Cities vowel shift, the Southern vowel shift, and the /ɑ/ ~ /ɔ/ merger.

In addition to Labov's acoustic measurement work on speech production, other researchers in the field of sociolinguistics have studied how this variation is perceived by naïve listeners. Preston's (1993) work on perceptual dialectology used several unique methods to investigate the kinds of mental representations college students have about dialect variation in the United States. In one study, he gave undergraduate students in Indiana, Hawaii, New York, and Michigan a map of the United States, including state boundaries, and asked them to indicate where people "speak differently." He found that most of his participants indicated some portion of the country as having a southern dialect and identified New York City as having its own unique accent. In addition, he found that participants tended to identify more regional varieties in close geographic proximity to their own hometown than in areas farther away, suggesting a gradient of knowledge about linguistic variation.

Preston (1993) also asked some of the students in this study to complete an attitude judgment task in which they were given a list of the 50 states and were asked to rate each state on the correctness, pleasantness, and intelligibility of the English spoken there. He found an overwhelming tendency for participants to indicate that the most correct English is spoken in northern and western states and that the

least correct English is spoken in southern states. Pleasantness ratings tended to be based on where the participants themselves were from, with their home state typically receiving a high pleasantness rating.

Perceptual dialectology studies provide interesting information about what kinds of representations naïve listeners have stored in memory about dialect variation and reveal important differences in these representations based on where the participants are from. However, these kinds of studies do not reveal the underlying psychological and linguistic processes that provide insights into how these listeners actually perceive or encode linguistic variation. Because the participants were not asked to listen to actual speech samples in making their responses, all of the data were based on linguistic representations stored in long-term memory instead of direct behavioral responses to speech stimuli.

A few dialect categorization studies have been conducted that ask naïve listeners to make direct behavioral responses to actual speech stimuli. Purnell, Idsardi, and Baugh (1999) conducted one study of dialect identification using the “matched-guise technique.” A single talker left answering machine messages for apartment landlords in various neighborhoods in the San Francisco area using three guises: African American Vernacular English (AAVE), Chicano English (CE), and Standard American English (SAE). Purnell et al. measured dialect identification by recording the number of phone calls that were returned for each guise in each neighborhood. They concluded that the landlords could in fact identify the racial dialect of the talker based on a short answering machine message because the number of returned phone calls for the SAE guise remained constant across all neighborhoods, while the number of returned phone calls for the AAVE and CE guises decreased with the minority population of the neighborhood.

In a perceptual study on regional dialect identification, Preston (1993) played short narratives spoken by nine middle-aged male talkers to naïve listeners in Michigan and Indiana. The talkers were from nine different cities on a north-south continuum from Dothan, Alabama to Saginaw, Michigan. The listeners heard each narrative passage once and were then asked to identify which city they thought each talker was from. In general, the listeners were able to make a coarse distinction between northern and southern talkers, although the boundary between north and south was slightly different for the two groups of listeners.

Finally, Williams, Garrett, and Coupland (1999) conducted a dialect categorization task on the regional varieties of the English spoken in Wales. They recorded two adolescent male talkers from each of six regions of Wales, as well as two adolescent male speakers of Received Pronunciation (RP). Williams et al. played these recordings back to naïve listeners in each of the six regions and asked them to categorize the talkers using an eight-alternative forced-choice task (the six regions in Wales, RP, and “don’t know”). Overall categorization performance was about 30% correct. In addition, the listeners were only able to correctly identify 45% of the talkers from their own region. Taken together, the results of these three dialect categorization studies suggest that naïve listeners are able to encode some information about dialect variation and they can use this information to identify where unfamiliar talkers are from. However, their performance is not perfect and in fact seems to be quite effortful.

Recent work in our lab at Indiana University has also explored how naïve listeners categorize unfamiliar talkers by dialect. In particular, Williams et al. (1999), Preston (1993), and Purnell et al. (1999) all suggested that listeners can use their knowledge of variation in their native language to identify where talkers are from. In order to investigate how naïve listeners of American English categorize talkers by dialect, we carried out a perceptual categorization study that was similar to Williams et al.’s earlier investigation using regional varieties of American English and native listeners of American English (Clopper & Pisoni, 2004b). All of the stimuli used in our experiments were drawn from a large corpus of read sentences. Our first research question assessed whether naïve listeners could reliably categorize talkers based on where the talkers were from.

Labov et al.'s (in press) work on dialect variation in the United States has shown the kinds of variation and variability we can expect to find in talkers from different regions of the United States. In order to evaluate which acoustic-phonetic properties listeners might be attending to in making their categorization judgments, we also conducted an acoustic analysis on two sentences that were read by all of our talkers (Clopper & Pisoni, 2004b). Thus, our second research question was designed to assess what acoustic-phonetic properties were available to the listeners and which ones they attended to in making responses in the categorization task.

Preston (1993) found differences between participant groups in all of his perceptual dialectology studies, suggesting that a participant's residential history may have a strong impact on how that person perceives and categorizes language variation. To study this problem, we conducted a second dialect categorization task using two groups of listeners that differed in their personal experience with dialect variation to see how residential history affected their performance on the task (Clopper & Pisoni, 2004a). Thus, our third research question focused on how prior linguistic experience affects performance in the dialect categorization task.

Finally, several recent studies in our laboratory by Nygaard et al. (1994) and Logan et al. (1991) have shown that short-term experience in the laboratory with spoken language variation affects performance on other language processing tasks. Therefore, we also conducted a perceptual learning study using the same perceptual categorization task to determine how categorization performance with unfamiliar talkers would be affected by prior laboratory training in the task.

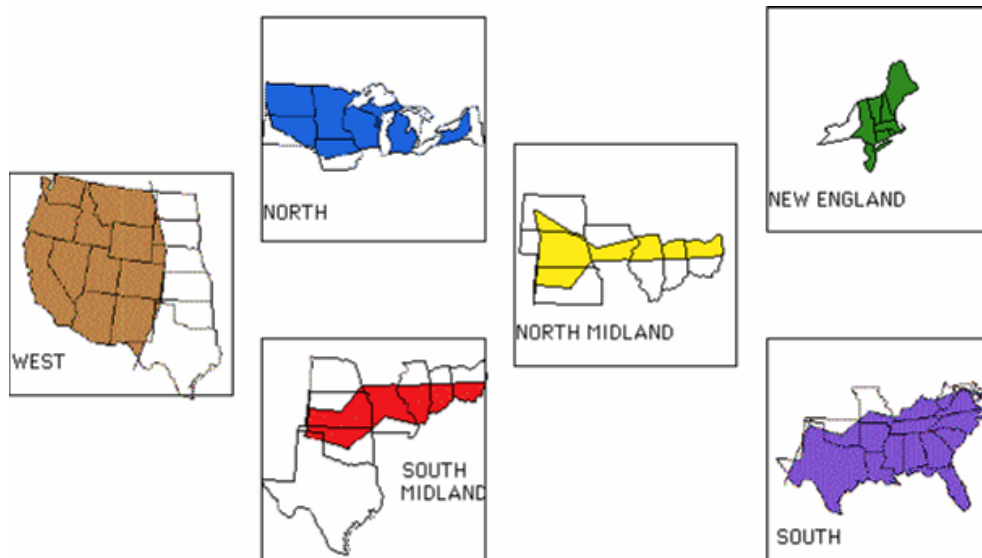
### **Experiment 1: Perceptual Categorization Task**

The purpose of the first experiment we conducted was to determine how well a group of naïve listeners can categorize talkers based on regional dialect of American English using a forced-choice categorization task. Utterances from sixty-six white male talkers in their twenties were selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986; Zue, Seneff, & Glass, 1990). The TIMIT corpus contains recordings of ten sentences read by 630 different talkers and was originally designed to contain a large degree of variability for use in speech recognition research. For our study, sentences were selected from eleven talkers from each of six different dialect regions in the United States: New England, North, North Midland, South Midland, South, and West.

Eighteen Indiana University undergraduates listened to sentences spoken by the 66 talkers. The listeners were divided post-hoc into three listener groups based on reported residential history: Northern Indiana (N = 7), Southern Indiana (N = 5), and Out-of-State (N = 6). The first two sentences that the listeners heard were those that were spoken by all of the talkers on the TIMIT corpus and are shown in (1). In addition, the listeners also heard each of the talkers reading a different, novel sentence. Examples of these novel sentences are shown in (2).

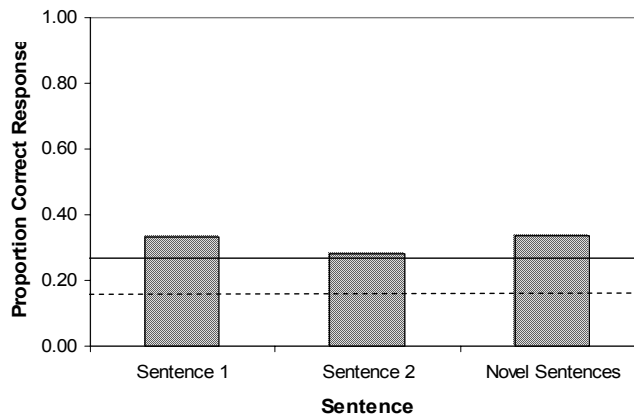
- (1)
  - a. She had your dark suit in greasy wash water all year.
  - b. Don't ask me to carry an oily rag like that.
  
- (2)
  - a. Beg that guard for one gallon of gas.
  - b. Barb's gold bracelet was a graduation present.
  - c. A huge tapestry hung in her hallway.
  - d. Clasp the screw in your left hand.

In the first phase of the experiment, the listeners heard all 66 talkers reading Sentence (1a) in random order and were asked to categorize the talker by dialect into one of the six geographic regions. The regions were presented on the screen as partial maps of the United States, including state boundaries, and were labeled with the name of the region, as shown in Figure 1. In the second phase of the experiment, the listeners heard all 66 talkers reading Sentence (1b) in random order and were again asked to categorize the talker by dialect. Finally, in the third phase, the listeners heard the 66 talkers reading the novel sentences in random order. No feedback was provided about the accuracy of their responses on this task.



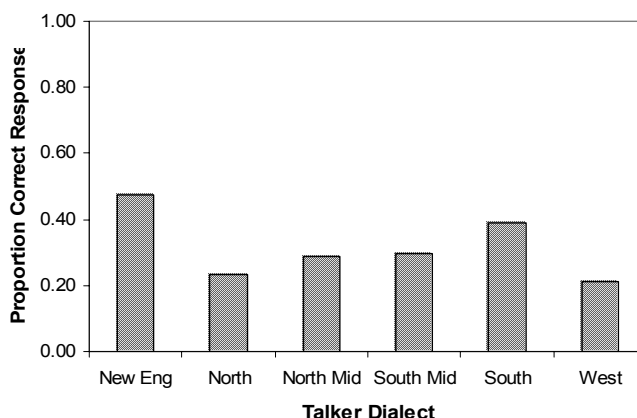
**Figure 1.** The six response alternatives in the categorization task. (From Clopper & Pisoni, 2004b).

The results expressed in terms of overall categorization accuracy are shown in Figure 2. The categorization scores revealed that the listeners performed quite poorly overall, although they were statistically above chance (17%) in all three phases of the experiment.



**Figure 2.** Overall proportion correct response in the six-alternative dialect categorization task, collapsed across all listeners and all talkers. Chance performance (17%) is indicated by the horizontal dashed line. Performance significantly above chance (25%), based on a binomial distribution, is indicated by the solid line. (Replotted from Clopper & Pisoni, 2004b).

Figure 3 shows the listeners' performance as a function of the six different dialect regions. The listeners were better able to correctly categorize the New England talkers than any other dialect group. They categorized the Southern talkers more accurately than the Northern or Western talkers. Unlike the talker differences, however, there were no listener group differences in any of the three phases and no individual listener differences within any of the three groups. Overall, the listeners performed consistently, although close to chance, on this task.



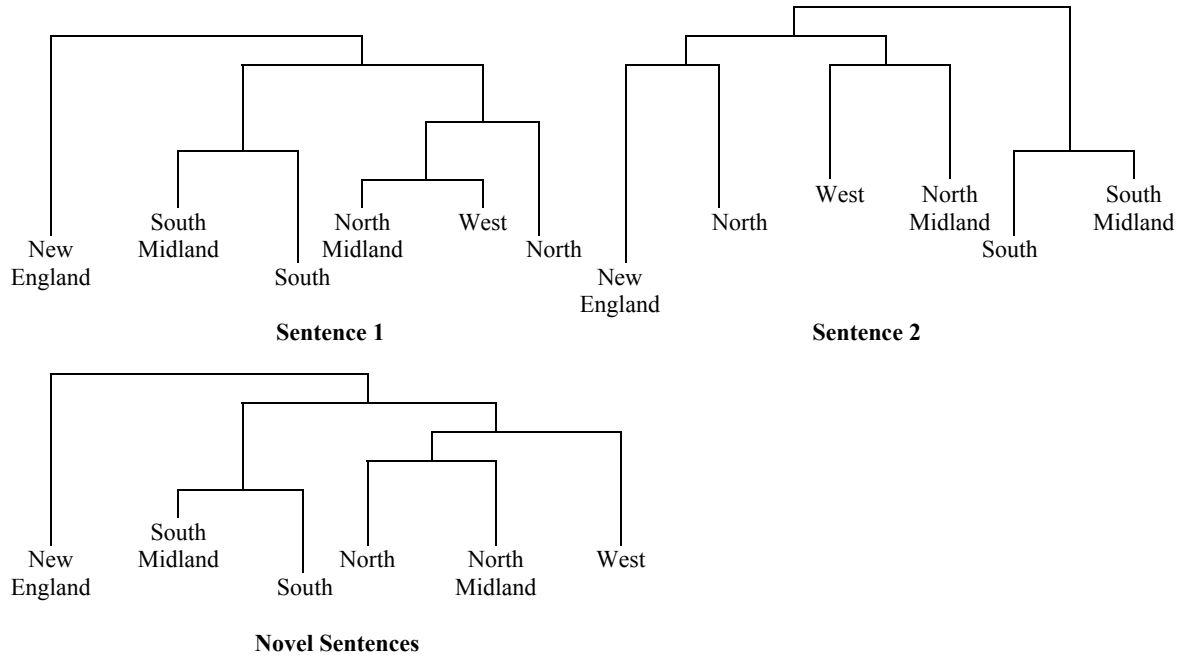
**Figure 3.** Proportion correct categorization responses for each of the six talker dialect regions, collapsed across the three sentence conditions. (Replotted from Clopper & Pisoni, 2004b).

In addition to the analysis of the correct categorization responses, we also conducted a clustering analysis using the confusion matrices of the error responses for each sentence. The confusion matrices were first submitted to the Similarity Choice Model (Nosofsky, 1985) to determine similarity and bias parameters. The similarity parameters indicate the degree of similarity between the dialect regions based on the listener errors. Examination of the bias parameters indicates response bias in selecting one response more often than another. The bias parameters revealed that the listeners' responses were relatively bias-free. The similarity parameters for each sentence were then submitted to ADDTREE (Cortier, 1995), which is an iterative additive clustering scheme that selects the most similar pair of cells in the matrix at each iteration to form a cluster and then recalculates the confusion matrix. The resulting clusters from this analysis are shown in Figure 4 for each of the three sentence conditions. In these figures, dissimilarity is indexed in terms of vertical distance in the display, so that the dissimilarity between any two dialect regions is the sum of the lengths of the least number of vertical lines connecting them.

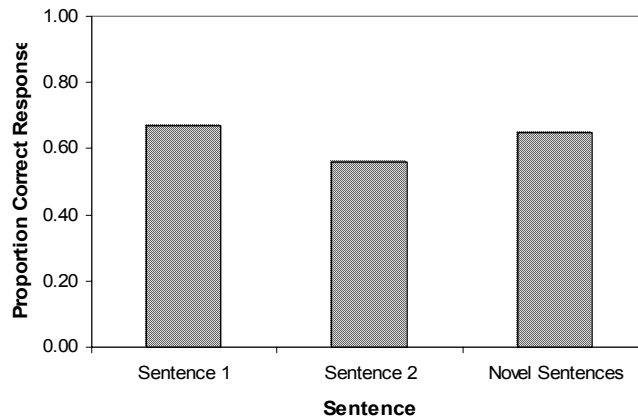
An examination of these clustering solutions revealed three main clusters of the categorization responses for each of the three sentences. For Sentence 1 and the Novel Sentences, the three main clusters were: New England; South and South Midland; and North, North Midland, and West. For Sentence 2, the three main clusters were: New England and North; South and South Midland; and North Midland and West. These results indicate that the confusions made by the listeners were not random, but instead suggest that the listeners were relying on three broad dialect regions in making their categorization judgments: New England, South, and North/West.

Categorization performance improves dramatically if it is measured in terms of the results of the clustering analysis. In particular, the responses on each sentence were rescored so that a response was scored as correct if it fell into the same major cluster as the stimulus item for that sentence. The results of this analysis are shown in Figure 5. This difference between the original measure of performance and the

new measure based on the confusions obtained in the categorization task provides additional behavioral evidence that the listeners were systematically making use of three dialect clusters instead of the six regions originally provided by the experimenters.



**Figure 4.** Clustering solutions for Sentence 1, Sentence 2, and Novel Sentences, collapsed across all listeners.



**Figure 5.** Proportion correct response in the categorization task when collapsed across the three main clusters for each of the three sentences. For Sentence 1 and Novel Sentences, collapsed across New England; South and South Midland; North, North Midland and West. For Sentence 2, collapsed across New England and North; South and South Midland; North and West.

Taken together, the results of this first perceptual experiment revealed that listeners are able to categorize talkers by regional dialect at performance levels above chance using a forced-choice task without feedback. In addition, the confusions made by the listeners were systematic in nature and enabled us to investigate patterns in their perception. Specifically, the results of the clustering analysis suggested that listeners make use of three broad perceptual categories instead of six: New England, South, and North/West. Thus, in answer to our first research question, we found that listeners can reliably categorize unfamiliar talkers by dialect using a forced-choice perceptual categorization task, although their performance is not error-free.

## Experiment 2: Acoustic-Phonetic Analysis

Our second experiment was designed to measure several selected acoustic-phonetic properties of these sentences and determine which ones listeners used in making their perceptual categorization judgments. Acoustic measurements were obtained from the first two sentences used in the previous experiment from all of the original 66 male talkers. The eleven acoustic measurements made for each talker are shown in Table 1.

Word	Segment	Measurement	Acoustic-Phonetic Property
dark	/a/	change in F3 (midpoint to offset)	r-fulness
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/	F2 (midpoint)	/u/ backness
don't	/ou/	F2 (midpoint)	/ou/ backness
		change in F2 (midpoint to offset)	/ou/ diphthongization
rag	/æ/	F2 (midpoint)	/æ/ backness
		change in F2 (onset to offset)	/æ/ diphthongization
like	/aɪ/	change in F2 (midpoint to offset)	/aɪ/ diphthongization
oily	/oɪ/	change in F2 (midpoint to offset)	/oɪ/ diphthongization

**Table 1.** Acoustic measures selected for comparison between dialect groups.

Each of the acoustic measures was expected to reveal differences between talker groups, based on what is known about phonological variation in American English (e.g., Labov et al., in press; Thomas, 2001). In particular, r-fulness (i.e., rhotic vs. non-rhotic) was predicted to distinguish the New England talkers from the other talkers, reflecting New England r-lessness. Vowel brightness in *wash* was predicted to reveal the *wash* ~ *warsh* alternation and to distinguish South Midland talkers from the others. Fricative voicing and fricative duration in *greasy* were predicted to distinguish the Southern and South Midland talkers from the others. In terms of vowels, /u/ backness was predicted to be lower for the Southern and Western talkers than for the other talkers. The degree of diphthongization of /aɪ/ and /oɪ/ was also

predicted to distinguish the South from the other regions. The degree of diphthongization of /ou/ as well as /ou/ and /æ/ backness were predicted to distinguish the North from the others. Finally, /æ/ diphthongization was predicted to distinguish New England from the others.

A summary of the overall results of the acoustic-phonetic analysis is provided in Table 2. New England talkers were significantly less rhotic than South Midland and Western talkers. Southern talkers had significantly greater voicing in the fricative in *greasy* than New England talkers and a significantly longer fricative in the same word than Northern talkers. South Midland, Southern, and Western talkers had significantly more fronted /u/'s than New England talkers. Northern talkers had a significantly centralized offglide in /ou/ relative to the Southern talkers and a significantly fronted /æ/ relative to New England talkers. Thus, several of the selected acoustic-phonetic measures revealed significant differences between the talker groups.

	New England	North	North Midland	South Midland	South	West
r-fulness ( $\Delta$ 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 ( $\Delta$ Hz)	-71	-148	-40	22	37	-41
/æ/ backness (Hz)	601	399	440	425	494	491
/æ/ diphthong ( $\Delta$ Hz)	256	177	255	280	223	233
/aɪ/ diphthong ( $\Delta$ Hz)	452	418	402	278	331	350
/oɪ/ diphthong ( $\Delta$ Hz)	301	384	434	250	226	445

**Table 2.** Acoustic measure means by talker dialect group.

In order to determine which acoustic-phonetic properties of the speech signal were good predictors of talker “dialect affiliation” (defined as the dialect labels used to classify the talkers in the TIMIT corpus) we conducted a series of logistic multiple regressions on the acoustic-phonetic measures and dialect affiliation. In each regression, the acoustic-phonetic measures were treated as potential predictor variables of dialect affiliation, which was scored dichotomously (“1” if the talker was from that region and “0” if he was not). The results of these analyses, shown in Table 3, revealed the acoustic-phonetic properties that were good predictors of actual dialect affiliation of the talkers (i.e., good predictors of the TIMIT labels). R-lessness and /æ/ backness were found to be good predictors of the New England talkers. For North dialect affiliation, /ou/ offglide centralization and monophthongal /æ/ were found to be good predictors. Fronting of /u/ and backing of /ou/ were good predictors of South Midland talkers. Finally, fricative voicing in *greasy* was a good predictor of Southern talkers. None of the acoustic-phonetic measures examined in this study turned out to be good predictors of either North Midland or Western talkers.

	<b>Significant Variables</b>	<b>Regression Coefficients</b>	<b>Overall <math>r^2</math></b>
<b>New England</b>	r-fulness	-.01	.33
	/æ/ backness	.02	
<b>North</b>	/ou/ diphthong	-.01	.21
	/æ/ diphthong	-.01	
<b>North Midland</b>	n/a		
<b>South Midland</b>	/u/ backness	-.01	.19
	/ou/ backness	.01	
<b>South</b>	fricative voicing	3.4	.21
<b>West</b>	n/a		

**Table 3.** Results of the logistic multiple regression analysis on acoustic-phonetic properties and talker dialect affiliation. For each of the dialect groups, the significant acoustic measures are shown with their regression coefficients and the overall  $r^2$  showing model fit. (From Clopper & Pisoni, 2004b).

	<b>Significant Variables</b>	<b>Regression Coefficients</b>	<b>Overall <math>r^2</math></b>
<b>New England</b>	r-fulness	-.36	.39
	/æ/ backness	.34	
	/ou/ diphthong	-.22	
	vowel brightness	.21	
<b>North</b>	/ou/ diphthong	-.38	.27
	/u/ backness	.29	
<b>North Midland</b>	/oi/ diphthong	.56	.31
<b>South Midland</b>	/u/ backness	-.26	.38
	vowel brightness	-.34	
	fricative voicing	.33	
<b>South</b>	/oi/ diphthong	-.39	.49
	/ou/ diphthong	.33	
	/u/ backness	-.33	
	/ou/ backness	.31	
	/æ/ diphthong	.20	
<b>West</b>	/oi/ diphthong	.40	.16

**Table 4.** Results of the linear multiple regression analysis on acoustic-phonetic properties and perceptual categorization. For each of the dialect groups, the significant acoustic measures are shown, along with their regression coefficients and the overall  $r^2$  showing model fit. (From Clopper & Pisoni, 2004b).

A second regression analysis was conducted to determine which acoustic-phonetic properties in the signal affected the listeners' categorization behavior in Experiment 1. In this set of linear multiple regressions, the acoustic-phonetic measurements were again treated as predictor variables and the categorization performance of the listener served as the dependent variable. The results of this analysis, summarized in Table 4, reveal those acoustic-phonetic properties that listeners were attending to in making their categorization judgments. In categorizing talkers as New England, listeners were attending to r-lessness, /æ/ backness, centralized /ou/ offglides, and vowel brightness in *wash*. For North, listeners were attending to centralized /ou/ offglides and backed /u/. Diphthongal /oi/ was a good predictor of identification of North Midland and Western talkers for these listeners. Fricative voicing in *greasy*, /u/ fronting, and a dark vowel in *wash* were all good predictors of categorization as South Midland. Finally, /oi/ monophthongization, /ou/ diphthongization, /u/ fronting, backed /ou/, and /æ/ diphthongization were good predictors of categorization as Southern.

Taken together, the results of this acoustic analysis revealed that the dialects of the talkers used in this study could be reliably distinguished based on several robust acoustic-phonetic properties in the speech signal. In addition, seven acoustic-phonetic properties were found to be good predictors of dialect affiliation in the first regression analysis. The second regression analysis revealed that the listeners were attending to 16 acoustic attributes of the talkers in making their categorization judgments. Of the seven predictors of dialect affiliation found in the first regression analysis and the 16 predictors of categorization behavior found in the second regression analysis, four overlapped: New England r-lessness, New England /æ/ backness, North /ou/ offglide centralization, and South Midland /u/ fronting. These results suggest that listeners are able to detect and perceive some of the acoustic-phonetic properties that distinguish talkers of different dialects and can use these properties reliably in responding in the categorization task. Thus, in answer to our second research question, we found that listeners used acoustic cues to r-fulness, vowel backness, and vowel diphthongization in making their categorization judgments.

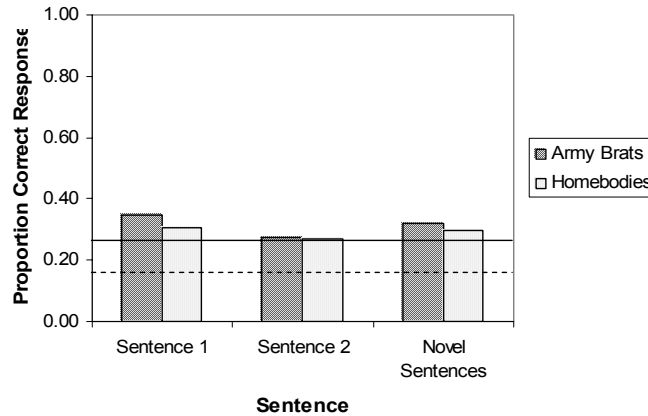
### **Experiment 3: Effects of Residential History: Army Brats vs. Homebodies**

Due to the post-hoc nature of the listener groups used in Experiment 1 and the small number of participants in each group, we failed to observe any systematic differences between the three groups in their categorization performance. However, given Preston's (1993) earlier findings, we expected to observe differences between listeners based on their past residential history. In addition, an extensive literature on language acquisition suggests that early linguistic experience and activities with many segmental, prosodic, and even indexical contrasts leads to better discrimination of those same contrasts later in life (e.g., Allen, 1983; Peng, Zebrowitz, & Lee, 1993; Polka, 1992; Strange, 1995; Tees & Werker, 1984). Based on these and other findings, early exposure to dialect variation might also be expected to have a lasting influence on a listener's perceptual abilities.

To explore this issue in greater depth, a third experiment was conducted to examine the effects of linguistic experience and residential history on listeners' performance in the dialect categorization task. The same set of sentence materials spoken by the same 66 talkers was used in this experiment as in Experiment 1. The experimental design was also identical to that used in the first experiment. Two new groups of Indiana University undergraduates participated as listeners in this study. The first group, the "Homebodies," consisted of 31 listeners who reported that they had lived exclusively in Indiana. The second group, the "Army Brats," consisted of 30 listeners who reported that they had lived in at least three states (including Indiana).

The perceptual categorization results of this study are shown in Figure 6. The Army Brats were more accurate overall on the six-alternative forced-choice categorization task than the Homebodies. These

results suggest that people who have lived in several different states perform better on the categorization task than people who have lived only in one state. Thus, greater linguistic experience and exposure to variation and variability through personal real-life interaction with people from different dialect regions leads to better performance on the dialect categorization task.



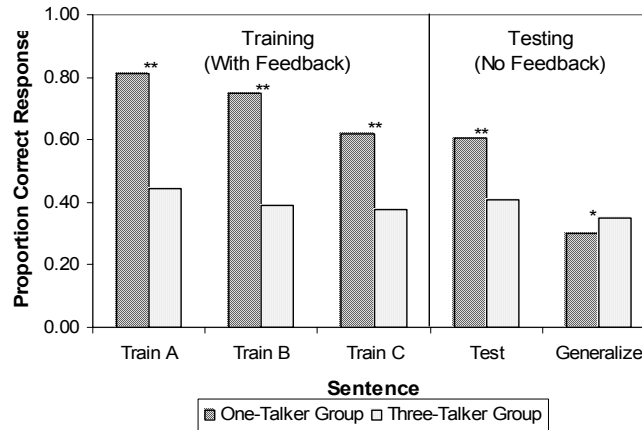
**Figure 6.** Proportion correct responses in each of the three phases of the dialect categorization task for the Army Brat listener group and the Homebodies listener group. Chance performance (17%) is indicated by the dashed line. Performance statistically above chance (25%) is indicated by the solid line. (Replotted from Clopper & Pisoni, 2004a).

#### Experiment 4: Some Effects of Perceptual Learning

The third experiment confirmed that real-life linguistic experience affects performance in the dialect categorization task. Experiment 4 was designed to investigate the effects of experience with linguistic variability in a laboratory setting on dialect categorization of unfamiliar talkers. The utterances from the same 66 talkers were used again in this study. Two additional groups of 30 Indiana University undergraduates participated as listeners. The first group, the “one-talker group,” was trained to identify one talker from each of the six dialects in three phases of training using the perceptual categorization task with feedback. The second group, the “three-talker group,” was trained to categorize three talkers from each of the six dialects in three phases of training. Following three blocks of training, the listeners in both groups were tested on the same talkers they had been trained on in the same categorization task without feedback to ensure that they had actually learned where the talkers were from. After this phase was completed, they participated in a generalization phase in which they heard unfamiliar talkers from each of the six dialect regions reading novel sentences and were asked to categorize them by dialect without feedback.

The results of this perceptual learning experiment are shown in Figure 7. While the group trained to identify one talker in each dialect performed better than the three-talker group on the three training blocks and the final test block using the same set of talkers, the group trained on three talkers actually performed better on the generalization phase with novel talkers than the one-talker group. This “cross-over effect” demonstrates that those listeners who were trained on materials with greater variability initially had more difficulty learning to categorize those materials, but were better able to generalize to new talkers in the final critical generalization phase. These perceptual learning results suggest that short-term exposure to greater stimulus variability even in a highly-controlled laboratory setting leads to better categorization of unfamiliar talkers. Thus, with regard to our third research question, greater linguistic

experience through laboratory perceptual learning also leads to better performance on the dialect categorization task.



**Figure 7.** Proportion correct responses in each phase of the perceptual learning experiment for each of the listener groups. \*\* $p < .001$ , \* $p < .05$  based on post-hoc t-tests.

## General Discussion

Experiment 1 revealed that naïve listeners are able to categorize talkers based on regional dialect with above-chance performance using a forced-choice perceptual categorization task. The clustering analysis of the response confusions suggested that listeners make use of three broad dialect categories: New England, South, and North/West. Acoustic analyses carried out in Experiment 2 revealed measurable acoustic-phonetic differences in production between the six dialect regions. In addition, multiple regression analyses suggested that listeners are able to use such stereotyped attributes as r-lessness and /ou/ centralization, as well as less stereotyped, but still prominent, attributes such as /u/ fronting and /æ/ backness to categorize talkers by regional dialect. In Experiment 3, we found that real-world exposure to talkers from different dialect regions as a result of living in a number of different geographic locations improves categorization performance. Finally, Experiment 4 revealed that short-term exposure in the laboratory to multiple talkers from each dialect region improves categorization performance on new talkers reading novel sentences, despite higher performance in training phases by the listeners who were exposed to only one talker from each region. Taken together, these results further confirmed that listeners can and do encode dialect variation in normal everyday language perception situations and that they can use the detailed knowledge they gain from these experiences in more formal laboratory tasks such as forced-choice dialect categorization.

The results of our new studies have several important theoretical implications for current models of speech perception and spoken language processing. Proponents of the traditional, abstractionist views of speech and language have stated that "... voice quality, speed of utterance, and other properties directly linked to the unique circumstances surrounding every utterance are discarded in the course of learning a new word" (Halle, 1985) and that "clearly most of the time anyone is listening to English being spoken, he [sic] is listening for the meaning of the message - not to how the message is being pronounced" (Brown, 1990). These traditional views of variation and variability are consistent with the generative linguistics paradigm in which language is described and modeled as being based on a one-to-one mapping between underlying phonological representations and surface phonetic forms. Generative approaches to

the study of language have recently even been applied to a theory of language evolution in early humans (Hauser, Chomsky, & Fitch, 2002). The results reported in this chapter, however, support the proposal that variation matters in speech perception and language processing and that listeners have access to fine acoustic-phonetic details and not merely symbolic representations of phonology (Pisoni, 1997).

The claim that variation matters is also supported by previous research on the role of talker variability in language processing tasks. As discussed earlier, Nygaard et al. (1994) showed that speech intelligibility in noise is better when the talkers are familiar than when they are unfamiliar. In addition, Mullennix et al. (1989) showed that listeners were better able to recognize words in noise when all of the words were spoken by a single talker than when the talker changed from trial to trial, suggesting that the listeners were sensitive not only to what was being said, but also to who was saying it. Mullennix and Pisoni (1990) also demonstrated interference in word and voice recognition tasks, revealing that listeners could not entirely ignore either the lexical content of the message or the talker, even when the perceptual task was to attend selectively to only one of the two dimensions. These findings are also consistent with “embodied” approaches to Cognitive Science which place the interactions of the body, mind, and the environment centrally in an understanding of cognition (Clark, 2001).

Taken together, these and other recent findings suggest that linguistic content and talker-specific indexical properties of spoken language are not separated in speech perception, but rather that they are both perceived, processed, and encoded as part of normal language processes. The results of Experiments 3 and 4 also suggest that exposure to linguistic variability, either in real life or in a laboratory setting, improves performance on dialect categorization, suggesting that listeners encode and store detailed acoustic-phonetic information about the dialect of talkers they are exposed to. These sources of information are not lost or discarded by the nervous system. The results of Experiment 2 suggest that listeners are successful in knowing what to listen for and attend to in making their judgments, without any explicit training or feedback. Models and descriptions of language in theoretical linguistics must therefore be able to account for the many-to-one mapping of surface forms to underlying forms that the listeners in these studies could make use of to identify where the talkers were from. In order to perform the categorization task in Experiments 1, 3, and 4 at levels above chance, the listeners had to access and explicitly use detailed phonetic knowledge that they have about variation in English surface forms to categorize the talkers by dialect. Although listeners are able to perform this task above chance, their performance was not perfect and many confusions were observed.

The present set of studies also show that the application of new experimental methodologies in the fields of Cognitive Psychology and Cognitive Science can be fruitfully applied to sociolinguistic issues to further our understanding of linguistic variation and how it is perceived and encoded by naïve listeners and how it is processed by the memory system. Future research on the perception of dialect variation might be able to make use of other methods to measure similarity such as free classification, paired comparison, and similarity scaling tasks. These tasks could provide further converging evidence for the basic underlying psycholinguistic processes used to identify and categorize dialects of English.

In our view, the process of speech perception involves not only the segmentation of the speech signal into meaningful linguistic units (e.g., words, sentences) and the recovery of the structure of the sound patterns, but also the processing and encoding of indexical information about the talker. This talker-specific information is available to the listener and can be used in laboratory tasks, such as categorizing unfamiliar talkers. Certainly, gender differences come to mind as an obvious distinction that we can make based on what we know about how males and females talk. Yet these perceptual abilities have been ignored in theoretical discussions of whether or not talker-specific information is encoded and used in speech perception and spoken word recognition. The results of the present set of experiments demonstrate that even talker-specific characteristics that involve more complex acoustic-phonetic

properties and that may be more difficult to perceive or encode, such as regional dialect, are also encoded by naïve listeners, stored in long-term memory, and used in a range of processing tasks.

## References

- Allen, G.D. (1983). Linguistic experience modifies lexical stress perception. *Journal of Child Language*, 10, 535-549.
- Brown, G. (1990). *Listening to spoken English*. (2<sup>nd</sup> ed.). New York: Longman.
- Clark, A. (2001). *Mindware*. New York: Oxford University Press.
- Clopper, C.G., & Pisoni, D.B. (2004a). Homebodies and Army Brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*, 16, 31-48.
- Clopper, C.G., & Pisoni, D.B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32, 111-140.
- Cortier, J.E. (1995). ADDTREE/P Program for Fitting Additive Trees.
- Fisher, W.M., Doddington, G.R., & Goudie-Marshall, K.M. (1986). The DARPA speech recognition research database: Specifications and status. *Proceedings of the DARPA Speech Recognition Workshop*, 93-99.
- Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, 102, 655-658.
- Halle, M. (1985). Speculations about the representation of words in memory. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 101-104). The Hague: Mouton.
- Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569-1579.
- Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099-3111.
- Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279-312.
- Klatt, D.H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169-226). Cambridge, MA: MIT Press.
- Labov, W. (1972). The social stratification of (r) in New York City department stores. In *Sociolinguistic patterns* (pp. 43-69). Philadelphia: University of Pennsylvania Press.
- Labov, W., Ash, S., & Boberg, C. (in press). *Atlas of North American English*. Mouton deGruyter.
- Lieberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /t/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89, 874-886.
- Mattingly, I. G., & Liberman, A. M. (1990). Speech and other auditory modules. In G.M. Edelman, W.E. Hall, & W.M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps* (pp. 501-520). New York: Wiley.
- Miller, G.A. (1946). Articulation testing methods. In *Transmission and Reception of Sounds Under Combat Conditions*. Summary Technical Report of Division 17, NDRC. Washington, DC. pp. 69-80.
- Mullennix, J.W., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, 47, 379-390.
- Mullennix, J.W., Pisoni, D.B., & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378.
- Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice-model analysis. *Perception and Psychophysics*, 38, 415-432.

- Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.
- Peng, Y., Zebrowitz, L.A., & Lee, H.K. (1993). The impact of cultural background and cross-cultural experience on impressions of American and Korean male speakers. *Journal of Cross-Cultural Psychology*, 24, 203-220.
- Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
- Pisoni, D.B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J.W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego: Academic Press.
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics*, 52, 37-52.
- Preston, D.R. (1993). Folk dialectology. In D. R. Preston (Ed.), *American dialect research* (pp. 333-378). Philadelphia: John Benjamins.
- Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18, 10-30.
- Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.
- Tees, R.C., & Werker, J.F. (1984). Perceptual flexibility: Maintenance or recovery of the ability of discriminate non-native speech sounds. *Canadian Journal of Psychology*, 38, 579-590.
- Thomas, E.R. (2001). *An acoustic analysis of vowel variation in New World English*. Durham, NC: Duke University Press.
- Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D.R. Preston (Ed.), *Handbook of perceptual dialectology* (pp. 345-358). Philadelphia: John Benjamins.
- Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication*, 9, 351-356.