RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 25 (2001-2002)
Indiana University

Perception of Dialect Variation: Some Implications for Current Research and Theory in Speech Perception[1]

Cynthia G. Clopper and David B. Pisoni

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 DC 00012 to Indiana University. We would also 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.

Perception of Dialect Variation: Some Implications for Current Research and Theory in Speech Perception

Abstract. Despite the mounting evidence that variation and variability play an important role in spoken language processing, few speech researchers have investigated the relationship between dialect variation and human speech perception. Sociolinguists, on the other hand, have extensively documented linguistic variation and its social implications, but have largely ignored how dialect variation is perceived and encoded by naïve listeners. We review and discuss several different methodologies that have been used to study the perception of dialect variation. Data collected from map drawing tasks in sociolinguistics, matched-guise studies in social psychology, caricatures in forensic linguistics, and perceptual categorization in cognitive psychology have all contributed to our understanding of how linguistic variation is perceived, processed, encoded, and used by naive listeners in normal language situations. The implications for these findings for models of speech perception, speech recognition and speech synthesis technologies, and theoretical linguistics are discussed.

Introduction

Variability in speech comes in many forms: within-speaker variability, cross-speaker variability, segment realization variability, and word environment variability (Klatt, 1989). One approach to the study of speech perception and spoken language processing is to ignore these sources of variability and to work to normalize the signal to find acoustic invariances across utterances, talkers, and contexts. A different approach, however, is to recognize these sources of variability as a natural consequence of language variation and work to understand how variation and variability are processed and encoded in speech perception. This second alternative espouses the notion that variation matters and that listeners can and do encode the indexical properties of the talker as part of the normal speech perception process (Pisoni, 1993).

Fifty years ago, Peterson and Barney (1952) recorded 33 men, 28 women, and 15 children reading two lists of ten [hVd] syllables. They took first and second formant frequency measurements for each of the vowels produced by each of the talkers. A scatterplot of the F1 values by the F2 values for each talker revealed a vowel space containing large overlapping ellipses for each of the ten vowels. In their discussion, Peterson and Barney pointed out the continuous nature of the vowel space; there are not obvious breaks in the data as one moves from one phoneme to another in the F1 x F2 plane. In addition, they noted that the distribution of tokens for a single phoneme represents the enormous variability with which any given vowel is produced across different talkers.

More recently, Hillenbrand, Getty, Clark, and Wheeler (1995) and Hagiwara (1997) have replicated Peterson and Barney's (1952) findings with respect to individual talker variation in terms of [hVd] formant frequency measures. Both of these more recent studies also found differences in mean formant values across their talkers compared to the formant values in the Peterson and Barney study. In particular, Hillenbrand et al. found a dramatic shift in the low vowels of their talkers, reflecting the Northern Cities vowel shift that has taken place in the last 50 years in urban areas in the northern United States. Hagiwara, on the other hand, found a dramatic shift in the back vowels, reflecting the southern California trend of back vowel fronting. These two sets of results suggest that researchers who are interested in the study of human speech perception should consider not only the effects of talker variability on vowel formants in production, but also the impact of regional dialect variation on vowel production and the implications of these differences in spoken language processing.

While this acoustic-phonetic research in the speech sciences has been carried out, sociolinguists have conducted extensive research on vowel systems in the United States. Labov, Ash, and Boberg (in press) recorded 700 individuals across the country as part of their telephone survey (TELSUR) project. Based on an acoustic analysis of the vowels contained in the utterances, they have mapped the major and minor regional dialects of American English. The resulting atlas provides evidence for the major vowel shift phenomena that are currently taking place in North American English, including the Northern Cities shift, the Southern shift, the low-back merger found in the west and upper midwest areas, and Canadian raising. In addition, Thomas (2001) used the individual vowel spaces of nearly 200 talkers in various locations around the country and of several ethnic backgrounds as the basis for his description of vocalic variation in North American English, including detailed discussions of the vowel systems of communities in Ohio, North Carolina, and Texas, as well as African American, Mexican American, and Native American varieties. Finally, many other researchers in the fields of sociolinguistics and dialect geography have conducted small-scale studies of the vowel systems of regions from Maine to Missouri to California. The combined result of this effort is the mounting evidence for an enormous amount of variation in speech production as a result of regional and ethnic boundaries.

Despite the obvious relationship between speech perception research and sociolinguistic research on variation in production, speech perception researchers and sociolinguists have been working in almost complete isolation from one another. Speech researchers are interested in discovering ways to understand and model how humans perceive, process, and encode language and are faced with questions about acoustic-phonetic invariance in the speech signal and the role of different types of variability in language processing. In addition, theoretical linguists have also been working under the assumption that language is a symbolic system with relatively fixed underlying representations. Variation at the phonetic level has not been considered relevant to understanding, modeling, or describing language under this symbolic view. Therefore, until recently, variation in speech was treated as a source of noise; that is, as an undesirable set of attributes that needed to be reduced or eliminated.

On the other hand, sociolinguists are interested in describing natural variation as it occurs on social, regional, and ethnic levels and are faced with questions about the social implications of variability such as stereotypes, prejudice, and language attitudes as they impact the classroom and the workplace. Until recently, however, the question of how variation in language is perceived, processed, and encoded by listeners in order to allow them to make social judgments based on speech had been largely ignored by both speech researchers and sociolinguists. In this paper, we discuss some of the progress that has been made over the last 15 years in addressing the relationship between speech perception and dialect variation, as well as the implications of this research for studies of human speech perception, speech recognition and synthesis technologies, and linguistic theory.

Where Speech Perception and Sociolinguistics Intersect

Researchers in the fields of sociolinguistics and speech perception have provided large amounts of evidence to support the notion that linguistic variation between talkers due to regional and ethnic differences is real and robust. What we know less about, however, is what naïve listeners know about these sources of variation. Sociolinguists have spent much of their time documenting the linguistic variation that exists (Labov et al., in press) and speech perception researchers have spent their time trying to reduce or eliminate the variation (or ignoring it altogether) (Johnson & Mullennix, 1996).

There are a handful of methodologies, however, that have been used to investigate the question of what naïve listeners know about ethnic and regional linguistic variation. Some of these methodologies stem from the social psychology literature, such as attitude judgments and the matched-guise technique. Others have been developed in the field of perceptual dialectology, such as map-drawing tasks and dialect consciousness studies. Still others stem from the forensic linguistics literature, such as accent imitation and caricature. Finally, more recently a few researchers have employed methods developed in cognitive

psychology to explore the perception of variation in discrimination, matching, identification, and categorization tasks.

Map Drawing. One of the more unique methodologies employed by sociolinguistic researchers interested in the mental representations of dialect variation is the map drawing task designed by Preston (1989). In this task, naïve participants are given a map of the United States (or Brazil or Japan) and are asked to draw and label the areas where "people speak differently." The results of these studies reveal that the cognitive maps that these participants have of dialect variation do not correspond to the dialect maps that are drawn by sociolinguists and dialect geographers. In fact, while most undergraduates in the United States will identify some portion of the country as "South" and most can reliably identify New York City as having its own accent, composite maps of groups of participants invariably have one or more regions that are not labeled at all. That is, unlike dialectologists, naïve participants in these studies appear to believe that some regions of the United States are accent-free. In addition, Preston (1986) had students in Indiana, Hawaii, New York, and Michigan complete this map drawing task and he found that where the students were from had an effect on how they drew the maps. In particular, the participants tended to label more dialect regions in close geographic proximity to themselves than farther away. This finding suggests that naïve listeners are sensitive to the variation that they hear through personal experience with and exposure to people from areas surrounding their hometown or state.

While this kind of task reveals something about the mental representations that naïve listeners have about dialect variation, the task itself is based on judgments made from memory that may be highly biased and unreliable. The underlying assumption of the map drawing research is that the participants have some kind of full-formed mental representation of what they think the speech of a certain region sounds like. The results of these studies may not be able to provide a full understanding of speech perception or dialect perception, however, because they are based on measures of memory, not perception. In order to address issues of speech perception and dialect variation, researchers need to obtain some kind of behavioral response to actual spoken language stimuli. For example, participants could be given a map of the United States and after listening to a short sample of speech, they would be asked to indicate on the map all of the places that the talker might be from. Such a perceptual categorization task would reveal not only the participants' perception of the speech sample under study, but also would provide information about how the participants mentally represent dialect regions, because they would be indicating on their map all of the places where they believe people talk in the same way as the talker in the stimulus item.

Attitude Judgments. In other research, Preston (1989) asked his participants to make judgments about the "correctness," "pleasantness," and intelligibility of the English spoken in each of the 50 states. In general, he found that participants rated their own speech as most intelligible and most pleasant, but he made their correctness ratings based on what seems to be a set of perceived notions about where Standard American English is spoken. Specifically, western and northern states were typically identified as having the most "correct" English by all participants, regardless of where they were from. Similarly, southern states were identified as having the least "correct" English, even by participants from southern Indiana, who speak a variety of southern American English. These findings reflect what Preston calls "linguistic insecurity." Participants who are linguistically secure with respect to the variety of English that they speak are more likely to label their own variety as "correct" than participants who are linguistically insecure.

Like the map drawing task above, however, these attitude judgments rely on participant reports that are based on mental representations of language and there is no evidence to suggest that the participants necessarily have personal experience with or first-hand knowledge of the varieties of English that they label as least pleasant or most correct. These judgments could instead be highly biased and based on social stereotypes found in the media or perceived norms taught in the classroom by prescriptive grammarians.

Matched-Guise Technique. Another methodology that is commonly used in studies of language attitudes, particularly with respect to ethnic and racial varieties, is the matched-guise technique (Lambert, Hodgson, Gardner, & Fillenbaum, 1960). In a matched-guise experiment, listeners hear utterances read by a single talker assuming multiple guises (e.g., dialects, varieties, or languages) and are asked to rate the talker on scales such as intelligence, friendliness, and socioeconomic status. By controlling the voice qualities of the talker by using only a single talker, researchers can be more confident that their results point to attitudes toward phonological properties of language varieties and not to inherent differences in voice quality between talkers of different varieties. Studies of this kind often find that nonstandard language varieties are rated lower than standard varieties on scales related to "intelligence" by all listeners, revealing a general tendency to relate linguistic standardness with intelligence (Linn & Pichè, 1982; Luhman, 1990). However, it is also often the case that speakers of nonstandard varieties will rate those varieties highly on scales related to "friendliness," showing solidarity with speakers of the same variety (Linn & Pichè, 1982; Luhman, 1990). These types of studies suggest that listeners can and do make a number of attitudinal judgments about a talker based on his or her speech and that in many cases, these judgments correspond to social stereotypes or prejudices often associated with the group that is represented by a certain language variety.

In these sociolinguistic and social psychology studies, there is no way to separate the attitude judgments made by the listeners from their ability to recognize the dialect of the speaker. The analysis of results collected using the matched-guise technique often assumes that the listeners first identified the racial, ethnic, or regional accent of the talker before making their attitudinal response. However, listeners in these tasks are rarely asked to identify where the speaker is from before (or after) making their ratings. It therefore seems premature to conclude from these studies that listeners think that speakers of Appalachian English, for example, are friendly and unintelligent when in fact the only conclusion that can be drawn is that when the talker is speaking in an Appalachian English guise, the listeners rate him or her as being friendly and unintelligent (Luhman, 1990). In addition, the issue of native-like performance in all of the guises used in this kind of study is often overlooked. The crucial assumption made in this research is that the talker is equally competent in all of the guises he or she uses. It is difficult to know to what extent the talker truly controls each dialect and to what extent the characteristic or stereotyped features of each dialect are merely caricatured.

Caricatures. A similar method that has been used in the forensic linguistics literature but that may also provide insight into what people know about language varieties is an imitation or caricaturization task. In one such study, Markham (1999) asked eight native speakers of Swedish to read a prepared passage and an unfamiliar passage using a number of different regional accents. He then asked linguistically-trained judges to listen to the passages and identify the accent as well as rate the reading on its naturalness and purity. Markham found that some talkers were indeed able to convincingly imitate some accents, even for native listeners of that accent. These results suggest that in some cases, listeners can not only perceive and represent the variation in the language around them, but they can also reproduce the phonological characteristics of non-native varieties accurately. By including both a prepared passage and a sight-reading passage, Markham was able to elicit several levels of dialect imitation productions.

An interesting follow-up and extension to this study would be to play the speech samples to untrained native listeners of the different varieties represented and ask them to identify where the talkers were from and rate the nativeness or naturalness of the productions. This kind of follow-up would permit an examination of what the naïve listener knows about his or her own variety, as well as provide another measure of the talkers' abilities to imitate and reproduce non-native varieties.

Dialect Consciousness. In a slightly different approach to the investigation of what listeners know about the linguistic features of varieties of their native language, Mase (1999) conducted what he called a "dialect consciousness" study. He asked a group of Japanese participants to list characteristics of Japanese dialects that they perceived as being different from their own. The participants were able to provide

grammatical, phonological, and lexical differences that distinguished their own dialect from the speech of the region in question. In addition, the features that the participants listed were typically those which are unique to a given region, and not those that are found in multiple dialects. That is, the participants were sensitive to the features that were characteristic of a single dialect as opposed to features that defined a broader region or group of dialects. Mase also studied the varieties actually spoken in the regions about which he had collected dialect consciousness data. He found that the characteristics provided by his participants were in large part quite accurate, although some of the properties tended to be older features that were used predominantly by the oldest generation or had died out completely, revealing a tendency for participants to report stereotypes that no longer reflected reality.

A parallel study has not been carried out in the United States, but a comparison of the results of the map-drawing task from American and Japanese participants suggests that Japanese speakers appear to have a better sense of the regional language varieties spoken in Japan than Americans do of American English varieties. It would be interesting to see to what degree native speakers of American English could correctly identify characteristics of southern speech or of a Boston accent. By investigating what features people think characterize a dialect, researchers may gain some insight into what features they are paying attention to when trying to determine where a person is from.

Phonological Description. Sociolinguists may not have spent their time asking naïve listeners to describe characteristic features of language varieties in the United States, but researchers such as Labov have devoted a great deal of effort to the description of variation in this country. As mentioned earlier, Labov et al. (in press) have compiled a large corpus of spoken language over the telephone and have analyzed the vowel productions of 700 speakers of American English. The results reveal several vowel shifts in progress. These data also provide a coherent account of the variation and variability in vowel production in the United States. While Labov's research is interesting and important from a documentation standpoint and while it provides researchers who are interested in variation with an excellent starting point for examining variation on a smaller scale in individual regions, states, or cities, the methodology is somewhat limited because it does not provide information about how this variation is actually perceived by naïve listeners.

Vowel Matching. One technique that does assess naïve listeners' perception of variation in production is the vowel-matching task used by Niedzielski (1999) in her study of the perception of the Northern Cities shift in Detroit English. In this task, listeners heard sentence-length utterances and were asked to select one of six synthesized vowel tokens that they thought matched the vowel in the target word in each sentence. Half of the listeners were told that the talker was from Detroit (as she actually was) and half of the listeners were told that the talker was from neighboring Canada. Niedzielski found that listeners who were told that the talker was from Canada most often selected the synthetic token that matched the actual vowel as the "best match." However, the listeners who were told that the talker was from Michigan most often selected the synthetic token that corresponded to a canonical (i.e., unshifted) vowel as the "best match." These results suggest that vowel perception is mediated by "knowledge" about the talker, such as where the listener believes the talker is from.

Niedzielski's (1999) conclusion was that Detroiters perceive themselves as speaking "standard" English, but that they perceive Canadians as speaking "with an accent" and this affected their perception of the vowels that they heard. One problem with this interpretation, however, lies in the design of the task. The listeners in Niedzielski's study were told to select the "best match" from six synthesized vowel tokens as part of a project on improving speech synthesis. It is possible and very likely that the group of listeners who were told that the talker was from Detroit selected canonical vowels because they wanted to be "helpful" to the experimenter by selecting the "best" vowels and not the "best match" vowels. In addition, although synthesized speech can be useful in tasks like this in which a range of tokens that are carefully controlled for formant values is necessary, synthetic speech samples are less natural than human speech productions and therefore research relying on behavioral responses to synthetic speech should be supplemented by converging evidence from studies involving natural speech.

Using a number of different methodologies from a variety of subfields of linguistics and psychology, researchers have begun to collect evidence to support the proposal that people can and do perceive and encode the variability in the speech they hear around them. Map-drawing, attitude judgment, and matched-guise tasks can provide researchers with valuable information about how people conceptualize the varieties of their native language. Caricature and dialect consciousness studies provide information about what the salient properties of a given language variety are and, in the case of caricatures, provide some insight into how well people can translate the knowledge they gain about linguistic variability through perception to production. Phonological studies of linguistic variation provide researchers with a basis for discussing what naïve listeners do and do not know about variation through thorough linguistic description. Finally, vowel-matching and other similar paradigms in cognitive psychology allow researchers to investigate perception of variation at a lower level of representation than the other kinds of tasks because, in ideal situations, they do not require the listeners to make more complex attitude judgments about the talkers.

The map drawing task and its associated attitude judgment tasks provide useful information about how naïve participants represent dialect variation. They help researchers answer questions such as: Where do people think people sound like each other? How many accents of American English do people think there are? What associations do people have with certain regional varieties? What do the mental maps of dialect variation look like for non-linguists? The matched-guise technique also provides some useful information about the kinds of attributes that naïve listeners associate with certain speech patterns. This methodology helps researchers to answer questions such as: What kinds of judgments do people make about people who talk a certain way? What kinds of attitudes towards people of certain races, ethnicities, socioeconomic status, and regional backgrounds can be elicited from listeners based on speech?

Caricature and dialect consciousness studies have focused on a slightly different aspect of perception, related to issues about naïve listeners' awareness of linguistic features. These kinds of studies help researchers answer questions such as: What do naïve listeners know about the linguistic features of other varieties? How accurate are listeners in describing or imitating characteristic features of other varieties? Can they imitate those features? Are their imitations good enough to fool native listeners? While these are all interesting questions, they are focused on the higher level of attitude representation, rather than perception.

Phonological descriptions of language varieties provide answers to questions about what the actual characteristic features of a dialect are. These descriptions can also help researchers answer questions such as: Where do people speak the same? How many dialects are there of American English? Which features are shared by multiple dialect groups? Which features are unique to a single region, ethnicity, or social class? Finally, the vowel-matching task provides some information about lower-level perception of variation. Niedzielski's (1999) study provided some initial answers to questions such as: How is perception influenced by a listener's beliefs about the utterance? To what extent does information beyond what is available in the acoustic signal impact perception?

The questions that remain to be investigated in dialect perception are those questions related to how listeners can actually use information in the speech signal to identify where a talker is from. Related issues include questions such as: What kinds of information does a listener encode with respect to dialect variation in everyday language situations? How is this information used in speech perception and language processing? How does linguistic experience with talkers from a variety of dialects affect a listener's ability to discriminate, identify, or describe different language varieties?

These kinds of questions can be explored using a wide variety of techniques developed in cognitive psychology, cognitive science, speech perception and spoken word recognition research. There are numerous experimental paradigms available in the speech perception literature that will allow researchers to investigate the perception of variation. For example, studies of dialect recognition or

categorization based on actual speech samples can provide new insights about what information about language variation is actually encoded in memory. Perceptual learning paradigms can be used to examine the role of linguistic experience in dialect identification, categorization, and discrimination. There has been some progress in this direction already over the last few years. The application of experimental methods to the study of linguistic variation should provide new insights into dialect perception and complement some of the earlier research using traditional sociolinguistic and social psychology methods.

Dialect Categorization

Dialect categorization studies are quite limited in the literature, but several researchers have developed methodologies to determine whether listeners can identify where a talker is from based only on a short speech sample. These perceptual studies employ traditional identification or categorization methodologies developed in the field of cognitive psychology for studying speech perception and spoken word recognition. Listeners hear short segments of speech spoken by a number of talkers and are simply asked to identify where they think the talker is from, using either a closed-set categorization task or an identification task. While these kinds of studies cannot answer questions about how the listeners use their knowledge of variation to make judgments about the talkers, they can provide new information about how listeners can use their knowledge of variation to determine where the talker is from. In combination with acoustic analyses of the speech signal and/or synthetic manipulation of the speech to highlight certain features, these kinds of studies can also be used to answer basic questions about which acoustic properties of the speech signal are most salient to listeners in identifying a talker's dialect. Through the study of relevant cues in dialect categorization, we can better determine what kinds of information about dialect variation are encoded, stored, and represented by the naïve listener based on his or her everyday experiences with linguistic variation in the environment.

Purnell, Idsardi, and Baugh (1999) conducted an implicit dialect identification experiment using the matched-guise technique. A single male talker using three racial guises (African American Vernacular English, Chicano English, and Standard American English) left answering machine messages for landlords in five neighborhoods in the San Francisco area. The researchers measured dialect identification by examining the relationship between the number of returned phone calls leading to appointments with a landlord from each neighborhood and the minority population living in each neighborhood. They found that the number of appointments for the Standard American English guise remained relatively constant across all five neighborhoods. However, the number of appointments for the African American Vernacular English and the Chicano English guises declined with the population of minorities in the neighborhood. Purnell et al. concluded that the landlords could identify the dialect, and therefore race, of the talker from just a brief sample of speech left on an answering machine.

Baugh (2000) has described the behavior of the landlords as "linguistic profiling" and has appeared on National Public Radio to discuss the findings of his study. While the issues related to racial identification are important, the original study itself was fundamentally flawed in several ways. The first flaw has to do with the matched-guise technique itself. As mentioned above, there are serious concerns about the ability of a single talker to produce utterances natively in multiple guises. The talker in the Purnell et al. (1999) study may not have had equally good control of all three guises. Second, the authors acknowledged that the dialects they used were "broadly" defined, but it is well known in the linguistic literature that variation among white speakers is much more regionally based than variation among African American speakers. Therefore, it is possible that similar results could be obtained using a northern white guise, a southern white guise, and a New York City guise. A relationship then might become apparent between perceived socioeconomic status and number of appointments made by the landlords that also corresponds to the mean socioeconomic status of a given neighborhood. While the results of this study do seem to show that people use their perception of dialect in making decisions in everyday life, the experiment itself did not control for the relationship between dialect and socioeconomic status.

In a more explicit study of dialect identification, Preston (1993) asked undergraduates in Michigan and Indiana to identify nine male talkers on a north-south continuum between Dothan, AL and Saginaw, MI. The talkers were all middle-age males and the speech samples were short utterances taken from longer narratives. The listeners heard each talker only once and were asked to identify which of the nine cities they thought he was from. While listeners were quite poor at identifying exactly where each talker was from, they were able to distinguish between north and south. The major boundary for the two groups of listeners was slightly different, suggesting that dialect identification is partly based on where the listener is from.

More recently, Preston (2002) has suggested that the difference in the location of the north-south boundary for the two listener groups could be related to differences in what they were listening for. In particular, his other studies have shown that Michiganders pride themselves on having the most "correct" English in the United States, while Hoosiers pride themselves on sounding "pleasant." Preston suggested that one possible explanation for the difference in perceived boundary in the identification task is that the Michiganders were making their identifications based on "correctness," while the Hoosiers were making their identifications based on "pleasantness."

One weakness of this study is that the listeners heard each talker only once and had to assign one talker to each city. Listeners therefore had to make their first response without reference to anything other than their own speech. They could make the remaining responses by comparing the voice on that trial with all of the voices they had heard previously. It has long been known in social and cognitive pscychology that behavioral responses to stimuli require reference and comparison to a standard. If a benchmark is not provided by the experimenter, then the participant must rely on his or her own internal standard which may shift in the course of the experiment (Helson, 1948). In order to reduce the effects of shifts in participants' standards for comparison, an alternative to this study might provide listeners with all nine talkers and the option to listen to each one as many times as they want and in any order that they want so that the listeners could each create their own continuum of the nine talkers, without being restricted to a single repetition of each talker presented in random order. Despite this methodological problem, Preston's (1993) study provides some additional evidence that naïve listeners can distinguish northern talkers from southern talkers. This research also gives some insight into what listeners might be doing to make these judgments, but we still do not know what specific acoustic properties of the speech signal listeners are basing their "correctness" or "pleasantness" judgments on.

The first study that explicitly investigated dialect categorization was conducted by Williams, Garrett, and Coupland (1999) on varieties of English spoken in Wales. They recorded two adolescent males from each of six regions in Wales and two speakers of Received Pronunciation (RP) telling personal narratives. The authors played short segments of these narratives back to different groups of adolescent boys from each of the six regions and asked them to categorize each talker into one of eight categories (the six regions of Wales, RP, or "don't know").

Overall, the listeners were able to correctly categorize the talkers with about 30% accuracy. Williams et al. (1999) also looked at the performance of each group of listeners on the two talkers from their own region and found that performance on same-dialect talkers was not much better than categorization performance overall. The average performance was about 45% correct on talkers from the same region as the listeners. While the talkers were selected from a larger set of recordings based on phonological criteria established by the authors, there was a significant difference in how well the two talkers from any given region were identified by the listeners (from the same region or from a different region). The authors suggested that this difference may be due to the availability of more or fewer salient phonological cues in some narratives or to the content of the narratives themselves as revealing something about the region in which the talker lived. While this study used spontaneous speech samples as its stimuli with the expectation of revealing the "true" dialect of the talkers, the lack of segmental and contextual control of the stimulus materials themselves does not allow us to consider what the perceptual differences between the talkers should be attributed to.

Some recent work by Clopper and Pisoni has also focused on the question of dialect categorization. In one set of studies, we considered the question of how well listeners could identify where talkers were from and what acoustic-phonetic properties of the speech signal the listeners might be using to categorize the talkers (Clopper & Pisoni, submitted). We selected sentence-length utterances from eleven male talkers in their twenties from each of six dialect regions in the United States from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Zue, Seneff, & Glass, 1990). Participants listened to the sentences and were then asked to categorize each talker into one of the six regions. In the first two phases of the experiment, listeners heard each of the talkers reading the same sentence. In the final phase of the experiment, listeners heard each of the talkers reading a different novel sentence.

Like Williams et al. (1999), we found that our listeners were only about 30% accurate in categorizing the talkers. Figure 1 shows the overall performance of the listeners in each of the three phases of the experiment. A clustering analysis on the confusion matrices of their responses revealed that listeners were not randomly guessing, but instead that they were making broad distinctions between New England, southern, and western talkers. As an example, the clustering solution for the sentence, "She had your dark suit in greasy wash water all year" is shown in Figure 2. In this representation of perceptual similarity, perceptual distance is represented by the lengths of the vertical branches. It should be noted that the three perceptual clusters roughly correspond to the three major regional dialects of American English that Labov and his colleagues have discussed in the phonological variation literature (Labov, 1998). Therefore, while overall performance was just above chance in terms of categorization accuracy, the results of the clustering analysis suggest that the listeners were responding in a systematic fashion and made categorization judgments based on three broader dialect clusters than those presented as response alternatives.

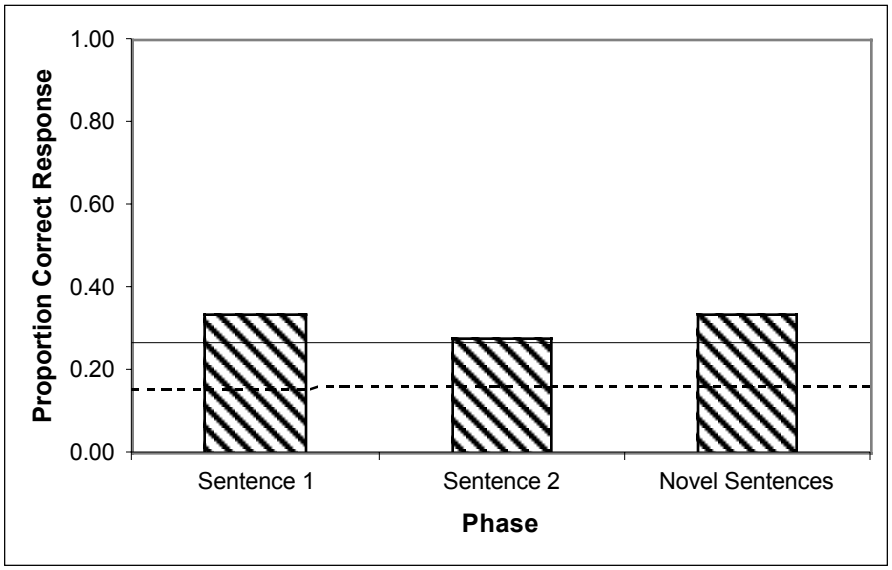

Figure 1. Proportion correct responses in each phase of the dialect categorization task. Chance performance (17%) is indicated by the dashed line. Performance statistically above chance (25%) is indicated by the solid line.

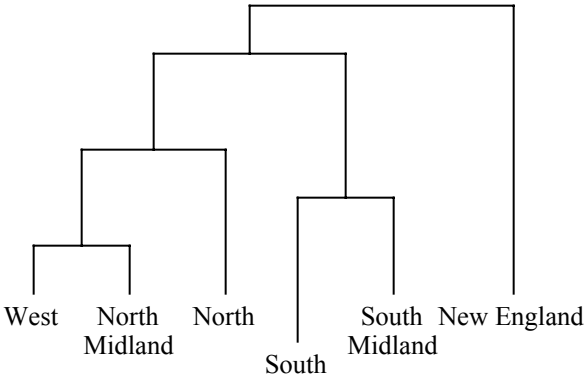

Figure 2. Clustering solution for the sentence, 'She had your dark suit in greasy wash water all year.' Perceptual distance is represented by the length of the vertical branches.

The sentence materials used in the categorization study were also subjected to an acoustic analysis. The acoustic measures confirmed that the talkers could be differentiated in terms of their dialect based on a number of reliable, well-defined acoustic-phonetic properties. In a logistic regression analysis, we found that there were seven acoustic-phonetic cues that were good predictors of dialect affiliation for our talkers. Table 1 shows the significant regression coefficients for the multiple regression analyses on the acoustic measures and dialect affiliation.

	Significant Variables	**Regression Coefficients**	**Overall r^2**
New England	r-fullness	-.01	.33
	/æ/ backness	.02	
North	/oʊ/ diphthong	-.01	.21
	/æ/ diphthong	-.01	
North Midland	n/a		
South Midland	/u/ backness	-.01	.19
	/oʊ/ backness	.01	
South	fricative voicing	3.4	.21
West	n/a		

Table 1. 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.

A similar regression analysis of the results of the categorization study with the results of the acoustic analysis revealed that listeners were attending to only four of the seven available cues in the speech signal. They were also attending to an additional 12 cues that were not good predictors of the dialect affiliation of these talkers. The results of this second set of analyses are shown in Table 2. The four overlapping cues revealed listeners' sensitivity to stereotypes (New England r-lessness and North /oʊ/ pronunciation) and to prominent but less stereotyped variations (New England /æ/ backing and South Midland /u/ fronting). The results of the categorization study and the acoustic analysis together suggest

that listeners can broadly categorize talkers by dialect and that they are able to make use of several reliable and robust cues in the speech signal to do so.

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

Table 2. Results of the linear multiple regression analysis on the acoustic-phonetic properties and perceptual categorization responses. For each of the dialect groups, the significant acoustic measures are shown with their regression coefficients and the overall r^2 showing model fit.

In a follow-up to the dialect categorization study, we investigated the role of the residential history of the listener on dialect categorization performance. In several studies, Preston (1989; 1993) has shown that participants from different parts of the country perform differently on his map-drawing and attitude judgment tasks. In our study, we asked two groups of listeners to carry out the same dialect categorization task described above. The first group ("homebodies") consisted entirely of listeners who had lived exclusively in Indiana. The second group ("army brats") consisted entirely of listeners who had lived in at least three states (including Indiana). We hypothesized that the listeners in the "army brat" group would perform better on the categorization task than the "homebodies" because through their real-life experiences living in a number of different places they would have been exposed to more variation than listeners who had lived exclusively in only one state.

Our results confirmed this hypothesis. The listeners in the "army brat" group performed better overall than the listeners in the "homebody" group. Figure 3 shows the proportion correct performance in each of the three phases for each of the listener groups. The clustering analysis on the data in this experiment also revealed differences in the perceptual similarity spaces of the dialects for the two listener groups, although the overall finding for both groups reflects the basic three-cluster structure (New England, South, West) found in the first experiment. These results confirm our intuition that personal experience with linguistic variation is an important contributing factor in how well people can identify where talkers are from based on their speech.

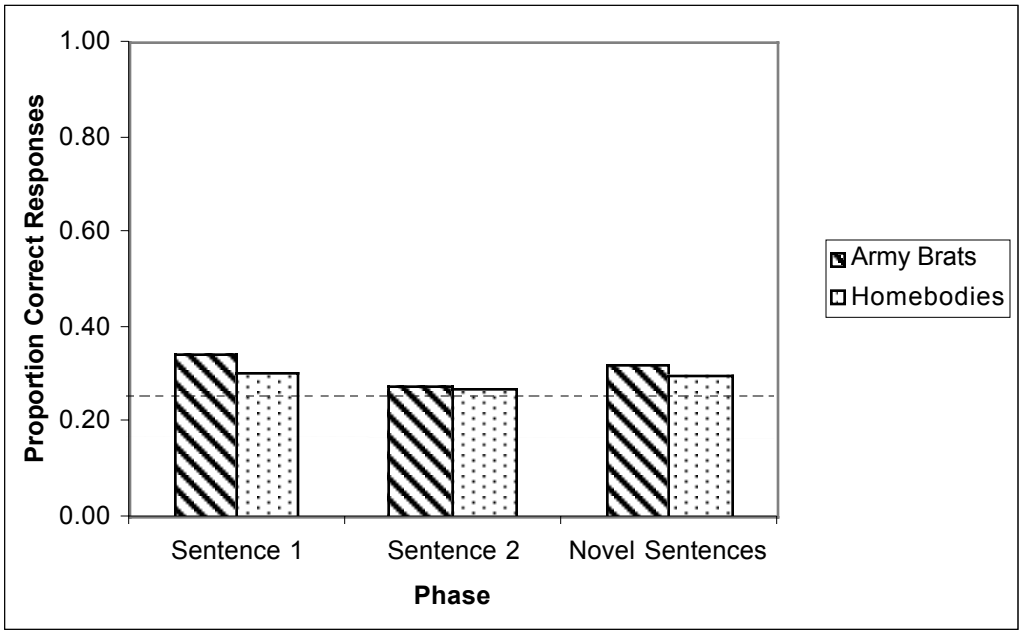

Figure 3. Proportion correct responses in each phase of the dialect categorization task for the Army Brat listener group and the Homebodies listener group. Chance performance (17%) is indicated by the dashed line. Performance statistically above chance (25%) is indicated by the solid line.

Perceptual Learning of Dialect Variation. Training and perceptual learning studies are often used in cognitive psychology to ensure that poor performance on a given task is not due merely to the participants' unfamiliarity with the task itself and to determine how much participants can improve and at what level their performance will asymptote (Green & Swets, 1966). Therefore, in order to determine whether or not personal experience in a laboratory setting would produce improvements in categorization performance, we conducted a set of short-term perceptual learning studies in which listeners were asked to learn to categorize a subset of the talkers used in the previous categorization tasks and then to generalize to new talkers. One group of listeners was trained to identify one talker from each of the six regions (the "one-talker" group). A second group of listeners was trained to identify three talkers from each of the regions (the "three-talker" group). Training consisted of three phases in which both groups of listeners heard sentences and were asked to categorize the talker by dialect. In the first two phases, the talkers all read the same sentence. In the third phase of training, every talker read a different, novel sentence. Feedback was given after every trial to aid in learning. Following the three training phases, the listeners participated in a test phase using the same talkers as in the training phases but without feedback to ensure that they had learned which talkers were from where. The last phase of the experiment was the generalization phase in which the listeners heard sentences read by all new talkers and were asked to categorize them without feedback. In both the test and generalization phases, the talkers all read different, novel sentences.

Categorization performance results are shown in Figure 4 for each of the five phases of the experiment for each of the two groups of listeners. While the one-talker group performed better in the training phases of the experiment, the three-talker group performed better in the generalization phase. This cross-over effect suggests that exposure to greater variation in training may produce more difficult initial learning in the training phases, but it results in better generalization to new talkers at test. Despite the fact that the training sessions for both groups were relatively short in comparison to other types of language-based perceptual learning experiments, listeners in the three-talker group were better able to categorize new talkers than listeners in the one-talker group. These results on perceptual learning of

dialect variation suggest that even when explicit instructions are not given about how to do the task, listeners know what to listen for and can extract information out of the acoustic signal that helps them in identifying the dialect of other unfamiliar talkers with very little exposure to the stimuli.

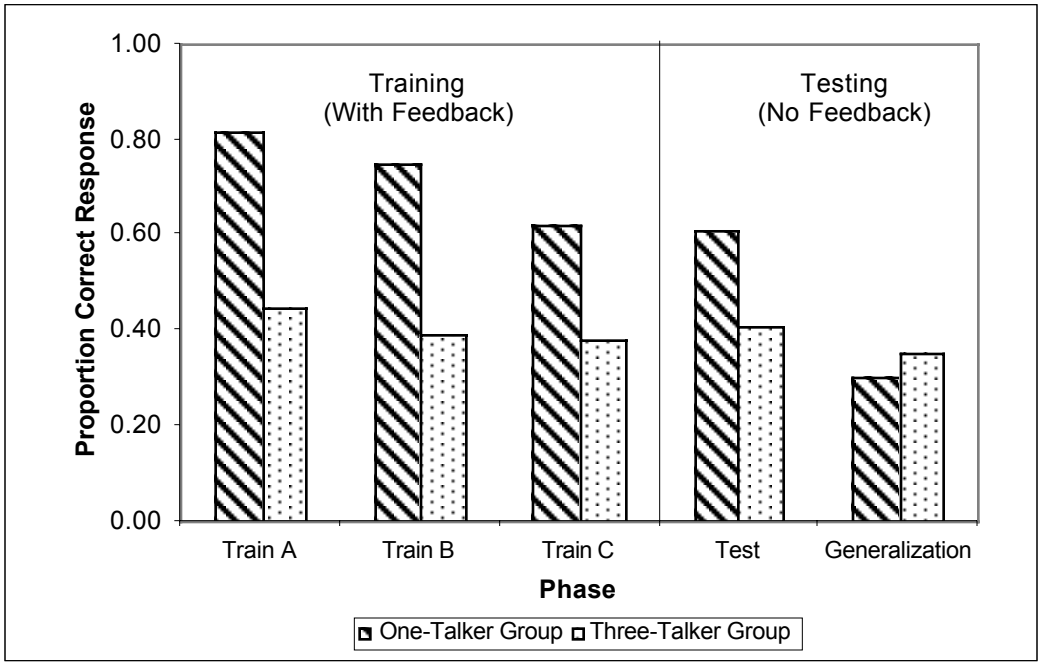

Figure 4. Proportion correct responses in each phase of the perceptual learning experiment for each of the listener groups.

The dialect categorization studies discussed above have a variety of goals with respect to the theoretical issues they wish to address. The matched-guise task focused on the judgments and decisions participants made based on their perception of where the talker was from (Purnell et al., 1999). The dialect categorization studies focused on how listeners made judgments about where a talker was from and what acoustic-phonetic properties of the speech signal the listeners were using to make such identifications (Clopper & Pisoni, submitted; Preston, 1993; Williams et al., 1999). Finally, the perceptual learning study examined the role that experience and learning have on dialect categorization abilities in naïve listeners. Despite these disparate goals, however, there is one theoretical claim that the results of all of the studies make that cannot be ignored: variation matters in the perception of spoken language. Naïve listeners can make reliable judgments about where an unfamiliar talker is from without explicit instructions about what to listen for. This ability suggests that listeners retain some kind of mental representation in memory of the varieties of their native language and that these representations develop naturally through a person's experience with and exposure to his community and the world at large. Specifically, recent findings from our lab have shown that greater personal experience and exposure with multiple dialects leads to better performance on the dialect categorization task. In addition, experience with greater variability of stimuli in perceptual learning paradigms of dialect variation also leads to better categorization performance on unfamiliar talkers. Therefore, experience both in real life and in the laboratory contribute to the information that listeners encode about the variation that they hear in the language around them.

Looking Forward

The relatively small literature investigating the relationship between dialect variation and speech perception in the laboratory means that there is still much work to be done before we can truly claim to understand how dialect variation is perceived, encoded, and represented in memory by naïve listeners. What little research has been done suggests that methodologies such as categorization tasks and perceptual learning tasks from the cognitive psychological literature and new methodologies developed by perceptual dialectologists such as map drawing tasks and the elicitation of dialect characteristics, as well as acoustic-phonetic analyses can provide converging information that will help us begin to answer fundamental questions about how listeners identify the dialect of a talker and how they use such categorizations once they have made them.

Methodological Extensions. There are several possibilities for extensions to the basic methodology of the dialect identification and categorization tasks. First, all of the major studies discussed above used only male talkers. This literature must be expanded to include studies of female speech and also studies in which the gender of the talkers varies from trial to trial. Sociolinguists have argued that women tend to be more conservative in their speech, often using fewer stigmatized forms (Labov, 1990). Speech stimuli recorded from male talkers might therefore be expected to reveal more regional or substratal forms. However, sociolinguists have also shown that women tend to be ahead of men in language changes in progress, regardless of whether the changes are above or below the level of conscious social awareness (Labov, 1990; Milroy & Milroy, 1993). Speech stimuli recorded from female talkers might therefore be expected to reveal current changes in progress. Niedzielski's (1999) perception work on Detroit speech suggests that listeners are not always aware of linguistic changes in progress, whereas Mase's (1999) dialect consciousness work suggests that people are aware of stigmatized forms that have become ingrained in popular social culture in the form of stereotypes. In order to understand the perception of both stigmatized and innovative forms, we need to extend the present research on dialect categorization to include responses to both male and female speech samples.

In addition, acoustic-phonetic research has traditionally involved only male talkers, due to the relative ease with which male formants can be measured as compared to female or child formants. However, recent acoustic analyses of male and female speech have shown that there are important acoustic differences between male and female speech in terms of segmental reductions (Byrd, 1994) and voice quality (Klatt & Klatt, 1990). Byrd's (1994) work also suggested an interaction between regional dialect and gender in segmental reduction that provides further support for the need to extend dialect categorization research to female talkers.

Second, the relatively poor performance of the listeners in the categorization tasks, their apparent ability to make broad categorical distinctions, and Preston's (1986; 1989) findings that naïve participants do not have cognitive maps that correspond to linguists' maps of dialect variation all suggest that in conducting these categorization tasks, we might want to reconsider the response format and alternatives that we provide for our listeners. Perhaps fewer response alternatives, representing the broad categories they show repeatedly in these tasks (e.g., North, South, and West) would result in better performance because it more directly reflects how listeners perceive and represent linguistic variation. Another alternative to the multiple choice tasks used in the studies discussed above would be a simple binary forced-choice distinction task in which listeners have to respond whether or not the talker has the same dialect as they do (e.g., "sounds like me" or "does not sound like me").

Third, all of the results described in this paper have relied on accuracy data in behavioral tasks. However, another common dependent measure used by psychologists interested in perception and processing is response latency. By adjusting the methodology slightly to force listeners to respond under time pressure, one could elicit response latency data that might also provide some insight into the underlying process of how listeners make their decisions. Are some varieties easier (faster) to identify than others? Are some listeners faster to respond than others? Do listeners respond faster to talkers from

their own region than to talkers from other regions? These are all questions that have not been investigated previously but that might provide further evidence about the role of variation in language processing, perception, and encoding.

The perceptual learning study discussed above also represents merely the tip of the iceberg of possible training methodologies that could be employed to investigate how listeners learn what to attend to. The one study that has been conducted so far involved short-term learning in a single session of less than one hour and short-term retention with the generalization phase immediately following the last training phase. While there was a small amount of improvement for the group who was trained on more talkers over the group who was trained on fewer talkers, neither group performed much above the levels of untrained listeners in other experiments. The question then remains at what level of performance the listeners would asymptote, if the training were continued over a number of sessions over a number of weeks. Similarly, the question also remains as to how long listeners would be able to retain whatever they had learned in the training sessions. Would the listeners exposed to more talkers still perform better on novel talkers after one day or one week?

In addition, the explicitness of the instructions could also be manipulated to determine whether or not asking listeners to focus on certain things affects their ability to learn to categorize talkers by dialect. For example, would the instruction to "focus on the vowels" cause listeners to improve even more over those without specific instructions, or do their strategies already include such a focus? Finally, like the categorization experiments above, the materials in both training and generalization phases need to be more varied to include not only talkers of both genders but also other kinds of utterances such as syllables, words, sentences, and perhaps even longer passages of connected speech.

Listener Populations. Another similarity between all of the categorization studies discussed above is that the talkers and listeners were all young to middle-aged normal hearing adults. It would be useful to extend this research to populations such as infants, children, older adults, non-native speakers, and hearing-impaired children and adults to investigate the effects of age, language background, and hearing impairment on dialect categorization. In particular, studies with infants and children would allow us to determine at what age the abilities to discriminate and categorize dialects arises. We might expect this ability to arise quite early in development, given some of the findings in the infant and child speech perception literature. For example, Houston and Jusczyk (2000) found that 10.5-month-old infants could separate the linguistic content of the speech signal from the indexical properties of the talker better than 7.5 month olds. Spence, Rollins, and Jerger (2002) have shown that 3-, 4-, and 5-year olds can use indexical information to identify cartoon characters by their voice. These findings suggest that indexical properties such as dialect variation are encoded in speech perception early in development and that children quickly learn to separate these talker-specific properties from the linguistic meaning.

Dialect categorization studies with older adult listeners would add to the discussion of the role of linguistic experience in dialect categorization. One hypothesis might be that older adults would perform better than younger adults because they have had more time to come into contact with more variation. In their study of dialect categorization in Wales, Williams et al. (1999) used two populations of listeners, adolescents and schoolteachers. Although they did not provide a statistical comparison of the performance between the two groups, the schoolteachers performed better (52% correct) than the adolescents (30% correct). Williams et al. concluded from these results that linguistic experience comes with age and that the difference between the two populations could be attributed to the greater experience of the teachers with linguistic variation. If performance continues to increase with age and experience, we might expect to find better categorization performance for older adults than for the college-aged listeners used in most of the studies discussed above.

When it comes to dialect categorization by non-native listeners, we might be inclined to predict that they would perform more poorly on a categorization task than native listeners. First, non-native listeners would typically have less experience with and exposure to the variation in the target language.

Second, they might be less sensitive to the variation in a second language than native speakers, particularly with respect to phonetic variation within a single phonological category. However, Bradlow and Pisoni (1999) found that non-native listeners were not more susceptible to talker variability effects in word recognition than native listeners, suggesting that some kinds of indexical variability have the same effects on all listeners. Dialect categorization research using non-native listeners would be an important contribution to our understanding of how linguistic variation is perceived and in what ways perception is constrained by language background.

Research involving hearing-impaired populations would provide evidence for the robustness of variation in cases where the signal is degraded. A case study conducted in our lab of a post-lingually deafened adult cochlear implant user on the perceptual learning task with training on a single talker from each region revealed poorer performance than normal hearing listeners in all of the training phases and performance at chance on the generalization phase. These results suggest that at least some of the information that is encoded by normal hearing listeners in this perceptual learning task is either not available to or not encoded by cochlear implant users. In addition, research on both adult and pediatric cochlear implant users has shown that they perform more slowly and less accurately on talker discrimination tasks than their normal hearing peers, suggesting that indexical information is not perceived and encoded in the same way for the two populations (Cleary, 2002; Kirk, Houston, Pisoni, Sprunger, & Kim-Lee, 2002). More research on these and other clinical populations would provide even further insight into the kinds of information that are available to and encoded by listeners in making these kinds of categorization judgments about language variation.

Other Measures of Perception. Another approach to the study of the perception of variation that has barely been examined is the notion of perceptual similarity spaces. The clustering analyses that we conducted on the confusion matrices from our dialect categorization studies reflect just one method of determining the perceptual similarity of the dialects we studied. Our results suggested that the perceptual similarity between dialects is based in part on the phonological similarity of the dialects, but that it also might be influenced by stereotyped uniqueness of a given variety. In particular, New England and South were often the most distinct dialects for our listeners and these two dialects are both associated with a number of stereotyped features. Other methodologies from the cognitive psychology literature such as paired comparisons, free classification, and similarity ratings tasks would provide converging evidence for the similarities between dialects (and between talkers within a given dialect) as they are perceived by naïve listeners.

While there is an extensive literature that outlines acoustic-phonetic differences between dialects in terms of vowel production, a complete discussion of the perception of variation in speech would also require a prosodic analysis of different language varieties. For example, are there consistent differences between dialects in terms of speaking rate (e.g., do southerners really talk more slowly?), fundamental frequency modulation, and stress? In addition to measuring these differences, it might be informative for researchers interested in the perception of dialect variation in the United States to conduct a dialect consciousness study like Mase's with American English speakers. A speaker's ability to articulate just what the characteristic aspects are of a given variety (his own or another's) will certainly reflect at least in part how he has represented that variety in memory. The results of these and many other possible studies will lead us to some better answers about how language variation is perceived, processed, stored, and used in human speech perception and spoken language processing.

Finally, electrophysiological and neuroimaging approaches to the study of the perception of dialect variation have barely been explored. Conrey (2001) reported the results of a vowel merger perception experiment in which she recorded reaction times in a cross-modal semantic priming task. She found that the behavioral reaction times in her study correlated with prior electrophysiology research on semantic priming. This correlation suggests that electrophysiology research on the perception of vowel mergers might also reveal interesting results that would provide further insights into the perception of variation. In addition, fMRI research has revealed some cortical distinctions between how linguistic form

and linguistic content are processed (Ni, Constable, Mencl, Pugh, Fulbright, Shaywitz, Shaywitz, Gore, & Shankweiler, 2000) and between first and second language processing in bilinguals (Kim, Relkin, Lee, & Hirsch, 1997). Future fMRI research on the processing of indexical variation and the role of linguistic experience with dialect variation in language processing may provide more information about how variation and variability are perceived, processed, and encoded by the human listener.

Implications for Speech Research, Speech Technology, and Theoretical Linguistics

There are many reasons why we need to gain a better understanding of dialect variation and perception. In terms of human speech perception, the more we know about how variation and variability are perceived, the better we will be able to understand and model spoken language processing. Many current models of speech perception assume that variation is stripped off early in a process of normalization so that the meaningful content of the signal can be recognized (Pisoni, 1997). This assumption is central to the traditional abstractionist view of speech and language as symbolic systems, in which the variation is treated as irrelevant noise. However, in order to completely understand the process of human speech perception, we need to understand how the sources of variability described by Klatt (1989) are perceived and encoded along with the linguistic message of the utterance. Researchers have only recently begun to abandon the traditional symbolic view of language and to investigate the contributions of linguistic variability in human speech perception. For example, in addition to the recent findings of the "army brat" study reported above, there is also an extensive literature on the role of talker variability and talker-specific information in speech perception that suggests that indexical properties of the talker are perceived and encoded by listeners in everyday linguistic tasks (e.g., Mullennix, Pisoni, & Martin, 1989; Nygaard, Sommers, & Pisoni, 1994). Dialect variation is clearly one of the indexical properties that is perceived and encoded in everyday language situations and its impact on speech perception deserves further investigation.

The implications for automatic speech recognition (ASR) systems with respect to variation and variability are perhaps even more striking. The variation and variability that exists in a single language is simply enormous and is constantly changing as the language changes. Human beings are able to adapt quickly to new talkers and linguistic changes, but ASR systems are still severely limited with respect to variation and change and require large amounts of training before they can accurately recognize speech. Ideally, ASR systems would be able to recognize not only a large number of lexical items, but also a large number of talkers and a large number of languages. However, most of the currently available commercial speech recognition systems are limited to a few talkers (e.g., personal computer speech-to-text software) or have limited vocabularies within a specialized domain (e.g., interactive automated flight information programs). One of the new areas of research in ASR systems is the "speech graffiti" project at Carnegie Mellon whose goal is to develop a universal speech interface that is more flexible than touchtone phone menu systems, but more rigid than a true natural language interface (Rosenfeld, Olsen, & Rudnicky, 2000). The idea behind the project is to build a human-machine speech interface that will be useful for an unlimited number of talkers across multiple domains, such as movie or apartment listings and flight information. The more we know about variation and how it is processed and encoded by human listeners, the more we will be able to apply our knowledge of human speech perception to building truly robust ASR systems.

Like ASR systems, speech synthesis technology is typically limited to a small number of voices and a limited vocabulary domain. The most natural synthetic speech can be built from the concatenation of resynthesized speech units smaller than the word, but larger than diphones. However, these systems are usually highly constrained in vocabulary. Successful speech synthesis of large vocabularies typically involves the concatenation of diphone strings, but the result is less natural speech (Black, 2002). Researchers at the University of Edinburgh in Scotland have been working to create a speech synthesizer using diphone concatenation that can produce speech in a number of different dialects of English, including Irish, Scottish, British, and American English varieties (Fitt & Isard, 1999). In addition to issues of prosody and sentence focus which remain problematic for speech synthesis programs (Wightman,

Syrdal, Stemmer, Conkie, & Beutnagel, 2000), "natural" speech synthesis must also be able to replicate important human features of speaking style such as register shifts and dialect variation, given the importance of such factors in human communication and interaction (Giles & Bourhis, 1976). As we learn more about what parts of the acoustic signal are important for human listeners in identifying where someone is from, we will be better equipped to design synthetic speech production systems that exhibit the appropriate characteristics of a given dialect.

Finally, research on the perception of dialect variation also has some important implications for theoretical linguistics. Like many speech perception researchers, theoretical linguists typically assume that each lexical item specifies one underlying phonemic input that is transformed through serial derivation or parallel candidate selection into a phonetic output. Generative phonologists typically assume a one-to-one mapping between phonemic forms in the mental lexicon and phonetic outputs in production. However, results of the studies on dialect caricatures and dialect consciousness discussed above suggest that naïve listeners have multiple mappings between underlying and surface forms, both productively and conceptually. In the sociolinguistics literature, variable rule analysis has been adopted by many researchers to account for variable phonetic outputs given a single underlying form in a single talker (Labov, 1969). However, acknowledging and accounting for the possibility of a one-to-many relationship between phonemic representations and phonetic forms in production has yet to occur in the mainstream generative paradigm. The research discussed above, however, suggests that phonological variation is important in human speech perception and any model of phonology would be remiss in overlooking this physically and psychologically real aspect of human language.

Cognitive scientists have recently begun to embrace again the notion of embodiment and explore the relationship between cognition and human interaction with the world (Núñez & Freeman, 1999). Recent work in the fields of speech perception and sociolinguistics crucially reveals that language is more complex than a simple symbolic system and that the perception of speech involves not only extraction of the linguistic meaning of the utterance, but also a number of other processes including identification of some of the indexical properties of the talker. Language as a cognitive process is therefore embedded in our physical and social interaction with the environment and any viable model of language processing must account for the variability inherent in actual language use.

Researchers in the fields of social psychology, sociolinguistics, forensic linguistics, psycholinguistics, and cognitive psychology have all contributed to the growing literature on the relationship between regional, social, and ethnic language variation and speech perception. The results of this diverse set of studies reveal that naïve listeners are aware of linguistic variation to the extent that they can imitate it, describe it, and use it to identify where people are from and to make judgments about social characteristics of the talkers. The implications of this research are widespread as well, including issues related to models of human speech perception, speech perception in clinical populations, child language development, speech recognition and speech synthesis technologies, neural biology, cognition and language, and theoretical linguistics. There is much work still to be done in this area, however, as well as a need for multi-disciplinary discussion of the results of these many and varied studies and the implications of these results for our understanding of human language.

References

Baugh, J. (2000). Racial identity by speech. *American Speech, 75*, 362-364.

Black, A.W. (2002). Text-to-speech synthesis. Paper presented at the 143rd Meeting of the Acoustical Society of America Short Course on Speech Technology and Conversational Systems. Pittsburgh, Pennsylvania, June 1-2.

Bradlow, A.R. & Pisoni, D.B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America, 106*, 2074-2085.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication, 15*, 39-54.
Cleary, M. (2002). Perception of talker similarity by normal-hearing children and hearing-impaired children with cochlear implants. Poster presented at the 143rd Meeting of the Acoustical Society of America. Pittsburgh, Pennsylvania, June 3-7.
Clopper, C.G. & Pisoni, D.B. (submitted). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*.
Conrey, B. (2001). Effects of dialect on merger perception. Poster presented at New Ways of Analyzing Variation 30. Raleigh, North Carolina, October 11-13.
Fitt, S. & Isard, S. (1999). Synthesis of regional English using a keyword lexicon. *Proceedings of Eurospeech 1999*, 823-826.
Giles, H. & Bourhis, R.Y. (1976). Methodological issues in dialect perception: Some social psychological perspectives. *Anthropological Linguistics, 18*, 294-304.
Green, D.M. & Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America, 102*, 655-658.
Helson, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review*, *55*, 297-313.
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.
Houston, D.M. & Jusczyk, P.W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1570-1582.
Johnson, K. & Mullennix, J. W. (Eds.). (1996). *Talker Variability in Speech Processing*. San Diego: Academic Press.
Kim, K.H.S., Relkin, N.R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature, 388*, 171-174.
Kirk, K.I., Houston, D.M., Pisoni, D.B., Sprunger, A.B., & Kim-Lee, Y. (2002). Talker discrimination and spoken word recognition by adults with cochlear implants. Poster presented at the 25th Mid-Winter Meeting of the Association for Research in Otolaryngology. St. Petersburg, Florida, January 26-31.
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.
Klatt, D.H. & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*, 820-857.
Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language, 45*, 715-762.
Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2,* 205-254.
Labov, W. (1998). The three English dialects. In M. D. Linn (Ed.), *Handbook of Dialects and Language Variation* (pp. 39-81). San Diego: Academic Press.
Labov, W., Ash, S., & Boberg, C. (in press). *Atlas of North American English*. Mouton deGruyter.
Lambert, W., Hodgson, E.R., Gardner, R.C., & Fillenbaum, S. (1960). Evaluation reactions to spoken languages. *Journal of Abnormal and Social Psychology, 60*, 44-51.
Linn, M.D. & Pichè, G. (1982). Black and white adolescent and preadolescent attitudes toward Black English. *Research in the Teaching of English, 16*, 53-69.
Luhman, R. (1990). Appalachian English stereotypes: Language attitudes in Kentucky. *Language in Society, 19*, 331-348.
Markham, D. (1999). Listeners and disguised voices: The imitation and perception of dialectal accent. *Forensic Linguistics, 6*, 289-299.
Mase, Y. (1999). On dialect consciousness: Dialect characteristics given by speakers. In D. R. Preston (Ed.), *Handbook of Perceptual Dialectology* (pp. 101-113). Philadelphia: John Benjamins.
Milroy, J. & Milroy, L. (1993). Mechanisms of change in urban dialects: The role of class, social network and gender. *International Journal of Applied Linguistics, 3*, 57-77.

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.

Ni, W., Constable, R.T., Mencl, W.E., Pugh, K.R., Fulbright, R.K., Shaywitz, S.E., Shaywitz, B.A., Gore, J.C., & Shankweiler, D. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience, 12*, 120-133.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*, 62-85.

Núñez, R. & Freeman, W.J. (Eds.). (1999). *Reclaiming Cognition: The Primacy of Action, Intention, and Emotion*. Bowling Green, OH: Imprint Academic.

Nygaard, L. ., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

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. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. *Speech Communication, 13*, 109-125.

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.

Preston, D.R. (1986). Five visions of America. *Language in Society, 15*, 221-240.

Preston, D.R. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics*. Providence, RI: Foris.

Preston, D.R. (1993). Folk dialectology. In D.R. Preston (Ed.), *American Dialect Research* (pp. 333-378). Philadelphia: John Benjamins.

Preston, D.R. (2002). The social interface in the perception and production of Japanese vowel devoicing: It's not just your brain that's connected to your ear. Paper presented at the 9th Biennial Rice University Symposium on Linguistics: Speech Perception in Context. Houston, TX, March 13-16.

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.

Rosenfeld, R., Olsen, D., & Rudnicky, A. (2000). A universal human-machine speech interface. *Technical Report CMU-CS-00-114*. Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.

Spence, M.J., Rollins, P.R., & Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech, Language, and Hearing Research, 45*, 214-222.

Thomas, E. R. (2001). *An Acoustic Analysis of Vowel Variation in New World English*. Durham, NC: Duke University Press.

Wightman, C.W., Syrdal, A.K., Stemmer, G., Conkie, A., & Beutnagel, M. (2000). Perceptually based automatic prosody labeling and prosodically enriched unit selection improve concatenative text-to-speech synthesis. *Proceedings of the International Conference on Spoken Language Processing*.

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.