RESEARCH ON SPOKEN LANGUAGE PROCESSING

Technical Report No. 12

August 31, 2004

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

Supported by:

Department of Health and Human Services
U.S. Public Health Service
National Institutes of Health
Research Grant DC00111
and
Training Grant DC00012

LINGUISTIC EXPERIENCE AND THE PERCEPTUAL CLASSIFICATION OF DIALECT VARIATION

Cynthia G. Clopper

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Linguistics and Program in Cognitive Science

Indiana University

August 2004

Acknowledgments

I arrived in Bloomington, Indiana almost exactly five years ago. I knew very little about graduate school and not too much more about linguistics. I would like to thank my advisors David Pisoni and Ken de Jong for their advice and support over the last five years; their doors were always open and they have helped me become a researcher, a teacher, and a scientist. David challenged me to explore an interesting new field of research and always expected a little more from me than I thought I could produce. In addition, he provided innumerable resources over the years that allowed me to conduct my research effectively and efficiently. Ken taught me how to teach, mostly by example, but also as a supervisor to my own teaching efforts. He also treated me like a colleague, which meant that I learned how to be one. I would also like to thank the other members of my committee: Julie Auger, Dan Dinnsen, and Rob Nosofsky. They always asked just the right questions to force me to clarify my thinking and place my work in the broader contexts of sociolinguistics, theoretical phonology, and cognitive science.

I would not have survived the last five years without the support of my friends and colleagues. My fellow SRLers (Caitlin, Rose, Brianna, Allyson, Miranda, Connie, Lorin, and Luis) provided a stimulating and fun work environment; my fellow linguistics and cognitive science graduate students (Sean, Judy, Carrie, Ellie, Maureen, and Kyoko), not to mention a few English graduate students (Emily, Chene, and Mat), were always available for a cup of coffee, a $2 movie, or happy hour; my friends in the 10:30 choir at St. Paul's Catholic Center (Jennifer, Mark, Mary Clare, Chuck, Patti, Maria, Lourdes, Dick, and Carl) provided spiritual support and a constant reminder that there is more to life than graduate school; and the Duke Goonies (Dave, Lisa, Clare, Derek, Laurel, Liz, Pookie, Matt, Kelley, Kelly, and Chris) could always be counted on for weekend getaway excuses (four weddings in five years!) and a good laugh.

Finally, I would like to thank my family for their support. My parents, Sam and Susan, never once asked me why I wanted to study linguistics or what I planned to do with a Ph.D. in linguistics. Instead, they have tried to understand what it is that I do on a daily basis because they know that I have found something that I am truly passionate about. My parents have always provided everything I ever needed or wanted and I am grateful for their generosity and love. My brothers, Ben and Eric, never got tired of asking me why I was still in college, but I would have been more worried if they hadn't been teasing me for the last five years and I appreciate their love and support as well.

This work was supported by NIH NIDCD T32 Training Grant DC00012 and NIH NIDCD R01 Research Grant DC00111 to Indiana University, as well as an Indiana University Chancellor's Fellowship for Graduate Study, the Acoustical Society of America Raymond H. Stetson Scholarship, and an Indiana University Summer Dissertation Award. I would also like to thank Luis Hernandez and David Parsons for their technical assistance in the development of the experimental protocols, Darla Sallee for her help with many administrative tasks, and the members of the Dialect Database group who were instrumental in the design of the Nationwide Speech Project corpus: Allyson Carter, Connie Clarke, Caitlin Dillon, Jimmy Harnsberger, Rebecca Herman, Luis Hernandez, and David Pisoni.

Abstract

The effects of linguistic experience on the perceptual classification of phonological dialect variation were investigated in a series of behavioral experiments with naïve listeners. A new digital speech corpus was collected which contains audio recordings of five male and five female talkers from each of six dialect regions in the United States (New England, Mid-Atlantic, North, Midland, South, and West). The speech materials recorded from each talker included isolated words, sentences, passages of connected text, and conversational speech. Acoustic analyses of the vowel systems of the talkers confirmed significant phonological variation due to regional dialect. Perceptual classification of dialect variation was assessed using sentence-length utterances from the new corpus with a six-alternative forced-choice categorization task and a free classification task. The independent variables examined in this study reflected the residential history of the listeners. In particular, two levels of the variables 'geographic mobility' and 'geographic location' were crossed to produce four different listener groups. For the mobility variable, listeners were either mobile (lived in more than one dialect region) or non-mobile (lived in only one dialect region). For the location variable, listeners came from either the Northern or the Midland dialect region of the United States. While residential history did not produce differences in overall accuracy in the forced-choice categorization task or classification strategy in the free classification task, residential history was found to affect the perceptual similarity of the six regional varieties examined. Clustering and multidimensional scaling analyses revealed that both geographic mobility and location help to shape the perceived similarity between geographically local dialects. In particular, the Northern listeners perceived a greater similarity between Northern and Midland talkers than the Midland listeners, while the Midland listeners perceived a greater similarity between the Midland and Southern talkers than the Northern listeners. In addition, the perceptual similarity spaces of the mobile listeners were less affected by geographic location than the similarity spaces of the non-mobile listeners. The perceptual classification results from this study contribute to the growing literature on the effects of linguistic experience on the perception of dialect variation by naïve listeners.

Linguistic Experience and the Perceptual Classification of Dialect Variation

Table of Contents

CHAPTER 1: PERCEPTION OF DIALECT VARIATION[1]

Introduction

Phonetic variability in speech comes in many forms. Within-speaker variability, cross-speaker variability, segment realization variability, and word environment variability, as well as numerous others, all lead to tremendous acoustic-phonetic variation in the production of phonemes (Klatt, 1989). Almost from the beginning, the traditional approach to the study of speech perception and spoken language processing has been to ignore these important sources of phonetic variability and to rely on abstract phonemic descriptions that are immune to variability across utterances, talkers, and contexts. A fundamentally different approach, however, is to recognize that these sources of variability are natural consequences of linguistic variation and investigate how variation and variability are processed in speech perception. Proponents of this second alternative claim that variation in speech matters in perception and that listeners can and do encode details of the indexical properties of the speech signal as a routine part of the normal speech perception process (Pisoni, 1993, 1997).

More than 50 years ago, in what is now considered to be a pioneering study in acoustic-phonetic speech science, Peterson and Barney (1952) recorded 33 men, 28 women, and 15 children reading two lists of 10 [hVd] syllables. They obtained first and second formant frequency measurements for each of the vowels produced by each of the talkers. A scatterplot of the formant frequency measurements revealed a vowel space in which each of the 10 vowels occupied a large area and overlapped in the space with other vowel categories. This distribution of tokens for a single vowel reflects the enormous variability with which any given vowel is produced across different talkers. In addition, this study revealed the continuous nature of the vowel space because there were no obvious breaks between adjacent vowel categories in the F1 x F2 space.

More recently, Hillenbrand, Getty, Clark, and Wheeler (1995) and Hagiwara (1997) 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 large differences in mean formant values across their talkers compared to the formant values in the Peterson and Barney (1952) study. In particular, Hillenbrand et al. (1995) found a dramatic shift in the low vowels of their talkers, reflecting the Northern Cities Chain Shift that has taken place over the last 50 years in urban areas in the northern United States. The Northern Cities Chain Shift is a clockwise rotation of the low and low-mid vowels; [æ] is raised and fronted, [ɛ] and [ʌ] are backed, [ɔ] is lowered and fronted, and [ɑ] is fronted. Hagiwara (1997), on the other hand, found fronting of the high back vowels, which is a common feature of southern California speech. These two newer sets of acoustic-phonetic measurements confirm the robust effects of talker variability on vowel formants in speech production and also demonstrate the impact of regional dialect variation on vowel production.

While this acoustic-phonetic research was being carried out in the laboratory by speech scientists, sociolinguists have been engaged in conducting extensive research on vowel systems in the United States. For example, Thomas (2001) plotted 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. He included 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. More recently, Labov, Ash, and Boberg (forthcoming) recorded more than 700 individuals across North America as part of their telephone

[1] Sections of this chapter will appear in Clopper, C.G., & Pisoni, D.B. (in press). Perception of dialect variation. In D.B. Pisoni & R.E. Remez (Eds.), *Handbook of Speech Perception*. Malden, MA: Blackwell.

survey (TELSUR) project. Based on an acoustic analysis of the vowels contained in the utterances, they have mapped the major and minor dialect regions of American English. The resulting atlas provides quantitative evidence for the major vowel shift phenomena that are currently taking place in North American English, including the Northern Cities Chain Shift, the Southern Vowel Shift, and the low-back merger found in the West and Midwest. In addition to these larger projects, many other researchers working in sociolinguistics and dialect geography have conducted small-scale studies of the vowel systems of regions from Maine to California. The combined results of these research efforts provide converging evidence for an enormous amount of variation in speech production as a result of regional and ethnic background.

Despite the obvious close relationship between speech perception research and sociolinguistic research on variation in speech production, speech perception researchers and sociolinguists have been working in almost complete isolation from one another. Speech researchers are typically interested in discovering ways to understand and model how humans perceive, process, and encode spoken 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 (Johnson & Mullennix, 1997). In addition, theoretical linguists have been working under the assumption that language can be modeled as an idealized symbolic system with relatively fixed underlying abstract phonological representations (Kenstowicz, 1994). Until recently, variation at the phonetic level has not been considered relevant to understanding, modeling, or describing language under this symbolic view. For almost 50 years, variation in speech was treated as a source of noise; that is, as a set of attributes that were irrelevant to extracting the underlying representations on which symbolic processes operated. As such, phonetic differences between talkers were treated as an undesirable set of attributes that needed to be reduced or eliminated in order to reveal the true underlying linguistic properties of the message (Pisoni, 1997).

In contrast to the typical abstract symbol-processing psycholinguistic approach, sociolinguists have described natural variation as it occurs as a result of social, regional, and ethnic differences and they have explored 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 samples had been largely ignored by both speech researchers and sociolinguists. In this chapter, we describe some of the progress that has been made over the last 15 years in addressing the relationship between speech perception and linguistic variation.

Where Speech Perception and Sociolinguistics Intersect

Researchers working 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 and is an important property of spoken language. We know less about what naïve listeners know about these sources of variation. While sociolinguists have spent much of their time documenting the linguistic variation that exists (e.g., Labov et al., forthcoming), speech perception researchers have devoted their time and effort to reducing or eliminating these natural sources of variability or simply ignoring them entirely (Johnson & Mullennix, 1997).

A small handful of research methodologies have been used to investigate naïve listeners' knowledge and representations of ethnic and regional linguistic variation. Some of these experimental methodologies stem from the social psychology literature, such as attitude judgments (Preston, 1989) and the matched-guise technique (Lambert, Hodgson, Gardner, & Fillenbaum, 1960). Other methods

have been developed in the field of perceptual dialectology, such as map-drawing tasks (Preston, 1986). Still others stem from the forensic linguistics literature, such as accent imitation (Markham, 1999). Finally, more recently several researchers have employed experimental methods developed in cognitive psychology and cognitive science to explore the perception of variation in identification and categorization tasks (e.g., Clopper & Pisoni, 2004b; Preston, 1993; Williams, Garrett, & Coupland, 1999).

Map-Drawing Tasks

One of the more unique methodologies used by sociolinguistic researchers interested in the mental representations of dialect variation is the map-drawing task designed by Preston (1986). In this task, naïve participants are given a map of the United States and asked to draw and label the areas where they think "people speak differently." The results of these studies have shown that the cognitive maps that these participants have of dialect variation do not correspond precisely to the dialect maps that are drawn by sociolinguists and dialect geographers. In fact, while most adults 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 unique accent, composite maps of groups of participants invariably look more like cultural maps of the United States than sociolinguistic maps (Preston, 1986).

In one study using this methodology, Preston (1986) asked adults in Indiana, Hawaii, New York, and Michigan to complete the map-drawing task. He found that where the participants were from had a substantial effect on how they drew the maps. In particular, the adults he studied tended to label more dialect regions in areas closer to themselves than in more distant geographic regions. These findings suggest that naïve listeners are sensitive to the variation in speech that they hear through personal experience and exposure to people from areas surrounding their hometown or state.

More recently, Tamasi (2003) used a variation of the map-drawing task to elicit mental representations of dialect differences from participants in Georgia and New Jersey. She gave naïve adults a stack of index cards with the state names written on them and asked the participants to sort the cards into piles based on how people speak in each state. Like Preston (1986), she found that naïve participants reliably identified salient regional varieties of American English, such as Southern and Northeastern varieties. Taken together, these perceptual dialectology studies suggest that naïve participants have mental representations of salient cultural regions in the United States, but that they may not have accurate or highly detailed representations of regional dialect variation.

Attitude Judgments

In other research on perceptual dialectology, 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 although participants rated their own speech as most intelligible and most pleasant, they made their correctness ratings based on what appeared to be stereotypes 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 (1989) called "linguistic insecurity," a phenomenon related to participants' perception of the "correctness" of their speech relative to some undefined standard or norm. Participants who were more likely to label their own variety as "correct" are considered to be linguistically secure, because they reported that they speak "correct" American English. Participants who did not view their own dialect as "correct" are linguistically insecure, because they reported that they speak a less "correct" version of English than people from other parts of the country.

Matched-Guise Technique

Another research methodology that has been used in studies of language attitudes, particularly with respect to ethnic and racial varieties, is the matched-guise technique (Lambert et al., 1960). In a matched-guise experiment, listeners hear utterances produced by a single talker who assumes multiple guises (e.g., dialects, varieties, or languages). Listeners are asked to rate the talker on subjective scales such as intelligence, friendliness, and socioeconomic status. By using a single talker to produce the speech samples, variation in voice quality that can be attributed to individual anatomy and physiology is controlled, although differences in voice quality that are functional may still be present in the stimulus materials. By using the matched-guise technique, researchers can be more confident that their results reflect attitudes toward phonological, lexical, and/or syntactic properties of language varieties and not inherent differences in voice quality between talkers of different varieties.

Matched-guise studies have typically found 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 to intelligence. However, it has also been found that speakers of nonstandard varieties will rate those varieties more highly on scales related to "friendliness," showing solidarity with speakers of the same variety (Linn & Pichè, 1982; Luhman, 1990). These types of ratings studies suggest that listeners can and do make a number of attitude 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 linguistic variety.

In one novel application of this approach, Purnell, Idsardi, and Baugh (1999) conducted a 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 decreased as the population of minorities in the neighborhood became smaller. Purnell et al. (1999) concluded that the landlords could identify the dialect, and therefore race, of the talker from just a brief sample of telephone-quality speech left on an answering machine.

Dialect Imitation

Dialect imitation tasks have been used in the forensic linguistics literature to assess dialect imitation performance by untrained talkers (Hollien, 2001). In one 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 each of the passages and identify the accent as well as rate the reading on its naturalness and purity. Markham (1999) found that some talkers were able to convincingly imitate some accents, even for native listeners of that accent. His results suggest that in some cases, listeners are able to perceive and represent the variation in the language around them, as well as accurately reproduce the phonological characteristics of non-native varieties.

Dialect Consciousness

In a 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 provided grammatical, phonological, and lexical properties that distinguished their own dialect from the speech of the region in question. In addition, the features that the participants listed were typically those that 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 (1999) also studied the varieties actually spoken in the relevant regions and 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 may not reflect current linguistic variation.

Vowel Perception

Another experimental technique that has been used to assess naïve listeners' perception of variation in production is the vowel-matching task developed by Niedzielski (1999) in her study of the perception of the Northern Cities Chain Shift in Detroit English. Listeners heard sentence-length utterances and were asked to select from a set of six synthesized vowel tokens the one that they thought was the best match to the vowel in the target word. 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 (1999) 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 not absolute and invariant but is mediated by "knowledge" about the talker, such as where the listener believes the talker is from (see also Ladefoged & Broadbent, 1957).

More recently, Rakerd and Plichta (2003) also examined the perception of the Northern Cities Chain Shift by listeners in Detroit and the Upper Peninsula of Michigan. They presented their listeners with sentence-length utterances and asked them to identify the vowel in the target words as [ɑ] or [æ]. Half of the sentences were produced by a talker from the Michigan Upper Peninsula, which does not participate in the Northern Cities Chain Shift. The other half of the stimulus materials were produced by a talker from Detroit and contained vowels consistent with the Northern Cities Chain Shift. Rakerd and Plichta (2003) found that the Detroit listeners adapted their responses to the talker and perceived the category boundary between [æ] and [ɑ] at a higher second formant frequency for the Detroit talker than for the Upper Peninsula talker. The listeners from the Upper Peninsula, however, exhibited no differences in perceived category boundary across the two talker dialects, suggesting that familiarity with a given dialect is necessary for accurate categorization of shifted vowels.

Similar results have been reported by Evans and Iverson (2004) for vowel differences between Northern and Southern varieties of British English. In their experiment, listeners heard sentence-length utterances produced in two different regional varieties of British English by a single male talker and were asked to match synthetic vowel tokens to target words. Evans and Iverson (2004) found that both native Southerners and Northerners who had moved to Southern Britain perceived quality differences between the Northern and Southern productions for some vowels, but not for others. Native Northern

listeners who had not lived in Southern Britain did not perceive any of the vowel quality differences presented in the experiment. Their findings suggest that residential history has an effect on naïve listeners' ability to perceptually match vowel tokens produced by talkers from a different dialect region.

Dialect Categorization

A number of researchers have recently begun to use experimental techniques developed in the fields of cognitive psychology and cognitive science, such as categorization and identification, to examine the perceptual classification of regional varieties of English and Dutch. In one study, Preston (1993) asked naïve adult listeners from Michigan and Indiana to identify nine male talkers on a North-South continuum between Dothan, Alabama and Saginaw, Michigan. The talkers were all middle-age males and the speech samples were short meaningful utterances taken from longer narratives. The listeners heard each talker only once and were asked to select which of the nine cities they thought the talker was from. While listeners were quite poor at identifying exactly where each talker was from, they were able to distinguish between Northern and Southern talkers. The geographic location of the major dialect boundary for the two groups of listeners was slightly different, however, suggesting that dialect identification is affected by where the listener is from.

More recently, Preston (2002) 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 previous studies showed that Michigan listeners pride themselves on having the most "correct" English in the United States, while Hoosiers pride themselves on sounding "pleasant" (Preston, 1993). Preston (2002) suggested that one possible explanation for the difference in perceived boundary in the identification task is that the Michigan listeners were making their identifications based on "correctness," while the listeners from Indiana were making their identifications based on "pleasantness."

One of the first forced-choice dialect categorization studies was conducted by Williams et al. (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 then 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"). No feedback was provided about the accuracy of their responses. 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 better than overall categorization performance. The average performance was about 45% correct on talkers that were from the same region as the listeners.

Van Bezooijen and her colleagues (Van Bezooijen & Gooskens, 1999; Van Bezooijen & Ytsma, 1999) have conducted similar dialect categorization research in the Netherlands and the United Kingdom. In the Netherlands, Van Bezooijen and Gooskens (1999) asked native Dutch listeners to identify the province of origin of three male talkers from each of four regional varieties of Dutch in a forced-choice categorization task using speech samples taken from interviews with an experimenter. The listeners were able to accurately categorize 40% of the male talkers. Van Bezooijen and Ytsma (1999) found similar results with female Dutch talkers using read speech passages. In the United Kingdom, Van Bezooijen and Gooskens (1999) reported that native British English listeners could identify the area of origin of male British English talkers with 52% accuracy.

More recently, a series of speech perception experiments conducted by Clopper and Pisoni (2004b) has also focused on the question of dialect categorization. In one set of studies, we investigated 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. We selected sentence-length utterances from 11 male talkers in their twenties from each of six dialect regions in the United States from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986). Participants listened to the sentences and were then asked to categorize each talker into one of the six geographic regions. No feedback was provided.

Like Williams et al. (1999), we found that our listeners were only about 30% accurate in categorizing the talkers in a six-alternative forced-choice categorization task. However, a clustering analysis on the stimulus-response confusions matrices from the categorization task revealed that listeners were not randomly guessing in making their categorization responses, but instead were making broad distinctions between New England, Southern, and Midwest/Western talkers. These three perceptual clusters are similar to the three major regional dialects of American English that Labov and his colleagues have discussed in the phonological variation literature (Labov, 1998). While overall performance was just above chance in terms of categorization accuracy, the results of the clustering analysis revealed 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.

All but one of the dialect categorization studies described so far have used only male talkers. However, 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 more advanced than 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 phonological changes in progress. Recently, Clopper, Conrey, and Pisoni (in press) replicated the earlier categorization and perceptual clustering results with a set of female talkers and a set of mixed male and female talkers. Our results confirmed that perceptual dialect categorization performance is robust across gender and presentation conditions.

Acoustic Cues to Perceptual Dialect Classification

The stimulus materials used in the initial categorization study by Clopper and Pisoni (2004b) were also subjected to an acoustic analysis. Acoustic measures of the sentences confirmed that the talkers could be differentiated in terms of their dialect based on a number of reliable, well-defined acoustic-phonetic properties. A logistic regression analysis revealed seven acoustic-phonetic cues that were good predictors of dialect affiliation for our talkers. A similar regression analysis of the results of the categorization study with the measures obtained from the acoustic analysis revealed that the naïve listeners in the categorization task were attending to only four of the seven available cues in the speech signal. These listeners were also attending to an additional 12 cues that were not good predictors of the dialect affiliation of these talkers. The four overlapping cues revealed listeners' sensitivity to stereotypes (New England r-lessness in the word *dark* and Northern /oʊ/ offglide centralization in the word *don't*) and to prominent but less stereotyped variables (New England /æ/ backing in the word *rag* and South Midland /u/ fronting in the word *suit*).

The relationship between dialect categorization and acoustic-phonetic properties of different American English dialects was also examined by Plichta and Preston (2003). They played synthesized tokens of the word *guide* that contained a vowel ranging from diphthongal [ɑy] to monophthongal [ɑ:]

to naïve listeners and asked them to assign each token to one of nine cities on a North-South continuum between Saginaw, Michigan and Dothan, Alabama. They found a significant relationship between the assignment of the stimulus items and their degree of diphthongization, with the most diphthongal token assigned most frequently to Saginaw, Michigan and less diphthongal tokens assigned to more Southern cities. Plichta and Preston (2003) also reported significant differences in geographic assignment between all of their stimulus items with a monotonic relationship between degree of diphthongization and degree of perceived Northern-ness. Taken together, the results of these two studies suggest strong relationships between perceived regional dialect affiliation and specific acoustic-phonetic properties in the speech signal.

Effects of Residential History on Dialect Categorization Performance

In his perceptual dialectology studies, Preston (1986, 1993) showed that participants from different parts of the country performed differently on his map-drawing and attitude judgment tasks. We therefore conducted a study to examine the effects of the residential history of the listeners on dialect categorization performance (Clopper & Pisoni, 2004a). We asked two groups of young adults to carry out the same six-alternative dialect categorization task described above. The first group consisted of listeners who had lived exclusively in Indiana (the "non-mobile" group). The second group (the "mobile" group) consisted of listeners who had lived in at least three different states (including Indiana) at the time of testing. We hypothesized that the listeners in the mobile group would perform better on the categorization task than the non-mobile listeners because through their real-life experiences living in a number of different places they would have been exposed to more phonetic and phonological variation than listeners who had lived in only one state.

The categorization results confirmed our prediction. The listeners in the mobile group performed slightly better overall than the listeners in the non-mobile group (Clopper & Pisoni, 2004a). In addition, residents of each dialect region more accurately categorized talkers from that region than non-residents, suggesting an effect of geographic location as well as mobility in the categorization task. A clustering analysis on the data in this experiment also revealed differences in the underlying perceptual similarity spaces of the dialects for the two listener groups. Although the overall finding for both groups reflected the basic three-cluster structure (New England, South, Midwest/West) found in the original categorization experiment reported by Clopper and Pisoni (2004b), the mobile listeners tended to perceive greater differences between geographically contiguous regions than the non-mobile listeners. These results replicate and extend Preston's (1986, 1993) earlier findings from the map-drawing task which showed that personal experience with linguistic variation affects naïve participants' mental representations of dialect variation.

Perceptual Learning of Dialect Variation

Training and perceptual learning studies are often used in the field of cognitive psychology to ensure that poor performance on a given task is not due merely to the participants' lack of familiarity and experience with the experimental procedures and to determine how much participants can improve and at what level their performance will asymptote (Green & Swets, 1966). 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 what they had learned about regional variation to a new set of talkers (Clopper & Pisoni, in press).

One group of listeners was trained to identify a single talker from each of the six dialect 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 two phases in which both groups of listeners heard sentences and were asked to categorize the talker by dialect. In the first phase, the talkers all produced the same sentence. In the second phase of training, every talker read a different, novel sentence. Participants were told the correct response after every trial to aid in learning. Following the two 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. Finally, the last phase of the experiment was the generalization phase in which the listeners heard novel sentences produced by new talkers and were asked to categorize them without feedback. In both the test and generalization phases, the talkers all produced different, novel sentences. Because the sentences varied across the different phases of the experiment, listeners had to rely on properties related to dialect and not individual sentences or specific talkers.

While the one-talker group performed better in the initial training phases of the experiment, the three-talker group performed better in the final generalization phase. This "cross-over effect" between training and generalization phases suggests that while exposure to greater variation in training may lead to more difficult learning in the initial training phases, these conditions led to better generalization to new talkers at final test. The training phases for both groups of listeners lasted less than 30 minutes, which means that this task only examined short-term retention following short-term learning. Despite this relatively brief exposure to the training materials in comparison to other types of language-based perceptual learning experiments, however, the listeners in the three-talker group were better able to categorize new talkers than the listeners in the one-talker group. These results regarding the 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 dialect-specific information from the acoustic signal that helps them in identifying the dialect of other unfamiliar talkers producing novel utterances with very little exposure to the training and test stimuli.

Summary

Using a number of different methodologies from a variety of subfields of linguistics and psychology, speech scientists and sociolinguists 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 provide researchers with valuable information about how listeners conceptualize the different regional, ethnic, and social varieties of their native language. Imitation and dialect consciousness studies provide additional information about the salient properties of a given linguistic variety and further insights into how well people can translate the knowledge they gain about linguistic variability through perception to production. Vowel perception tasks and perceptual categorization experiments allow researchers to investigate the perception of variation at the level of linguistic and sociolinguistic categories.

Taken together, the body of research on the perception of dialect variation leads to one overwhelming conclusion: naïve normal-hearing listeners can make reliable judgments about where an unfamiliar talker is from without explicit instructions about what to listen for. This perceptual ability suggests that listeners retain a 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 to multiple dialects leads to better performance on the dialect categorization task. Experience both in real life and in the laboratory contributes to the information that listeners encode about the variation that they hear in the language around them.

New Directions

The relatively small literature investigating the relationship between dialect variation and speech perception in the laboratory means that there is still much to be done before we can fully understand how dialect variation is perceived, encoded, and represented in memory by naïve listeners. The initial research that has been done on these problems suggests that experimental methodologies such as categorization paradigms and perceptual learning tasks from cognitive psychology and new methodologies developed by perceptual dialectologists such as map-drawing tasks and the elicitation of dialect characteristics, combined with acoustic-phonetic analyses, can provide converging measures that will help us begin to answer fundamental questions about how listeners identify the dialect of a talker and how they use this knowledge in a range of speech perception and spoken language processing tasks.

The research reported in this dissertation expands on the previous perceptual dialect categorization work by Clopper and colleagues (Clopper et al., in press; Clopper & Pisoni, 2004a, b) in several important ways. First, a new speech corpus was developed specifically for use in acoustic-phonetic and perceptual analyses of dialect variation in the United States. The corpus contains audio recordings of male and female talkers representing six different regional varieties of American English producing a range of utterances from isolated read words and sentences to conversational speech. This new corpus more accurately reflects current sociolinguistic descriptions of dialect variation in the United States (e.g., Labov et al., forthcoming) than the TIMIT corpus, which makes it an important new tool in the study of the perception of regional dialect variation. Second, a new methodology, free classification, was introduced to further explore the perceptual classification of regional dialect variation. The free classification task allowed us to explore the role of verbal labels in perceptual classification behavior by naïve listeners by reducing the inherent geographic category structure in the task. Finally, residential history was investigated more systematically than in the previous study through the explicit manipulation of the listeners' geographic mobility and location.

Chapter 2 describes the new corpus of spoken language that was developed specifically for perceptual experiments related to regional dialect variation in the United States. Chapters 3 and 4 present the results of two perceptual classification experiments using stimulus materials from the new corpus. A six-alternative forced-choice categorization task was conducted as a replication of the previous research using the TIMIT corpus and is described in Chapter 3. Chapter 4 reports the results of an experiment using the novel methodology, free classification, to study the perceptual classification of dialect variation. The residential history of the listeners was manipulated in both sets of experiments to allow for further investigation of the roles of geographic mobility and location in the perception and representation of dialect variation. Chapter 5 provides a discussion of the theoretical implications of this work for sociolinguistics, speech science, and theoretical phonology.

References

Clopper, C.G., Conrey, B.L., & Pisoni, D.B. (in press). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology.*

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.

Clopper, C.G., & Pisoni, D.B. (in press). Effects of talker variability on perceptual learning of dialects. *Language and Speech.*

Evans, B.G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115*, 352-361.

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

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.

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.

Hollien, H. (2001). *Forensic Voice Identification*. San Diego: Academic Press.

Johnson, K., & Mullennix, J.W. (Eds.). (1997). *Talker Variability in Speech Processing*. San Diego: Academic Press.

Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Cambridge, MA: Blackwell.

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. (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 dialects of English. In M.D. Linn (Ed.), *Handbook of Dialects and Language Variation* (pp. 39-81). San Diego: Academic Press.

Labov, W., Ash, S., & Boberg, C. (forthcoming). *Atlas of North American English*. New York: Mouton de Gruyter.

Ladefoged, P., & Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*, 98-104.

Lambert, W., Hodgson, E.R., Gardner, R.C., & Fillenbaum, S. (1960). Evaluational 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. 101-114). 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.

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

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.

Plichta, B., & Preston, D.R. (2003). The /ay/s have it: Stereotype, perception, and region. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

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.

Rakerd, B., & Plichta, B. (2003). More on perceptions of /ɑ/ fronting. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

Tamasi, S.L. (2003). *Cognitive Patterns of Linguistic Perceptions.* Doctoral dissertation, University of Georgia, Athens.

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

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology, 18,* 31-48.

Van Bezooijen, R., & Ytsma, J. (1999). Accents of Dutch: Personality impression, divergence, and identifiability. *Belgian Journal of Linguistics, 13,* 105-129.

Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D.R. Preston (Ed.), *Handbook of Perceptual Dialectology* (pp. 345-358). Philadelphia: John Benjamins.

CHAPTER 2: THE NATIONWIDE SPEECH PROJECT CORPUS[2]

Introduction

Phonological Dialect Variation in the United States

Researchers have been documenting regional linguistic variation in the United States for more than a century. The American Dialect Society was founded in 1889 with the goal of collecting a comprehensive American English Dictionary. Krapp (1925) documented regional varieties of American English based on grammars and pronunciation guides dating back to the 18th century. Based on this research, he identified three main dialects of American English: Eastern, Southern, and Western (or General). Thirty years later, McDavid (1958) described the early Linguistic Atlas projects in the United States which documented lexical and phonological variation. McDavid (1958) concluded that the major dialects of American English are Northern, Midland, and Southern. He also acknowledged that these dialects are more sharply distinguished on the Atlantic seaboard and that more transition areas between dialects are found as one moves westward across the country. Carver (1987) also examined regional lexical variation but he described only two primary dialects of American English: Northern and Southern.

More recently, linguistic variation has been explored using acoustic-phonetic analysis techniques. Thomas (2001) provided acoustic-phonetic vowel spaces for nearly 200 individual talkers. Although Thomas (2001) made no explicit claims about specific dialect regions, he did group his talkers into a Northern group and a Southern group. The most comprehensive study of regional variation in American English was conducted by Labov and his colleagues (Labov, Ash, & Boberg, forthcoming). The Telephone Survey project at the University of Pennsylvania includes telephone interviews with 700 talkers representing all major urban areas in the United States. The recordings have been analyzed acoustically and Labov (1998) defined three major dialects of American English based on the vowel systems of his 700 talkers: Northern, Southern, and the "Third Dialect." This third dialect includes Eastern and Western New England, Western Pennsylvania (centered on Pittsburgh), the Midland, the West, and Canada. Labov (1998) described the Mid-Atlantic metropolitan areas from New York City to Washington, D.C. as "exceptions" to the three-dialect division because speakers from this region do not exhibit the characteristic properties of any of the three major dialects. Figure 1 displays a map of United States showing these regions, as defined by Labov and his colleagues. No data are available for the gray areas on the map because these regions are sparsely populated and the Telephone Survey project focused on 145 urban areas with an average population of 1.7 million people, ranging from 88,000 in Aberdeen, South Dakota to 17.6 million in New York City (Ash, n.d.).

The vowel system of the Northern dialect of American English is characterized by the Northern Cities Chain Shift (Labov, 1998). The Northern Cities Chain Shift is a clockwise shift of the low vowels that includes the fronting and raising of [æ], the fronting of [ɑ], the lowering of [ɔ], and the backing of [ʌ] and [ɛ]. [ɪ] is also reported to be backed in the Northern dialect as a parallel shift to [ɛ] backing. Figure 2 depicts the major features of the Northern Cities Chain Shift.

---

[2] I would like to acknowledge the contributions of Allyson Carter, Connie Clarke, Caitlin Dillon, Jimmy Harnsberger, Rebecca Herman, Luis Hernandez, and David Pisoni in the development of the Nationwide Speech Project corpus, including the compilation of the materials, the selection of the equipment, and pilot testing of both equipment and participants.

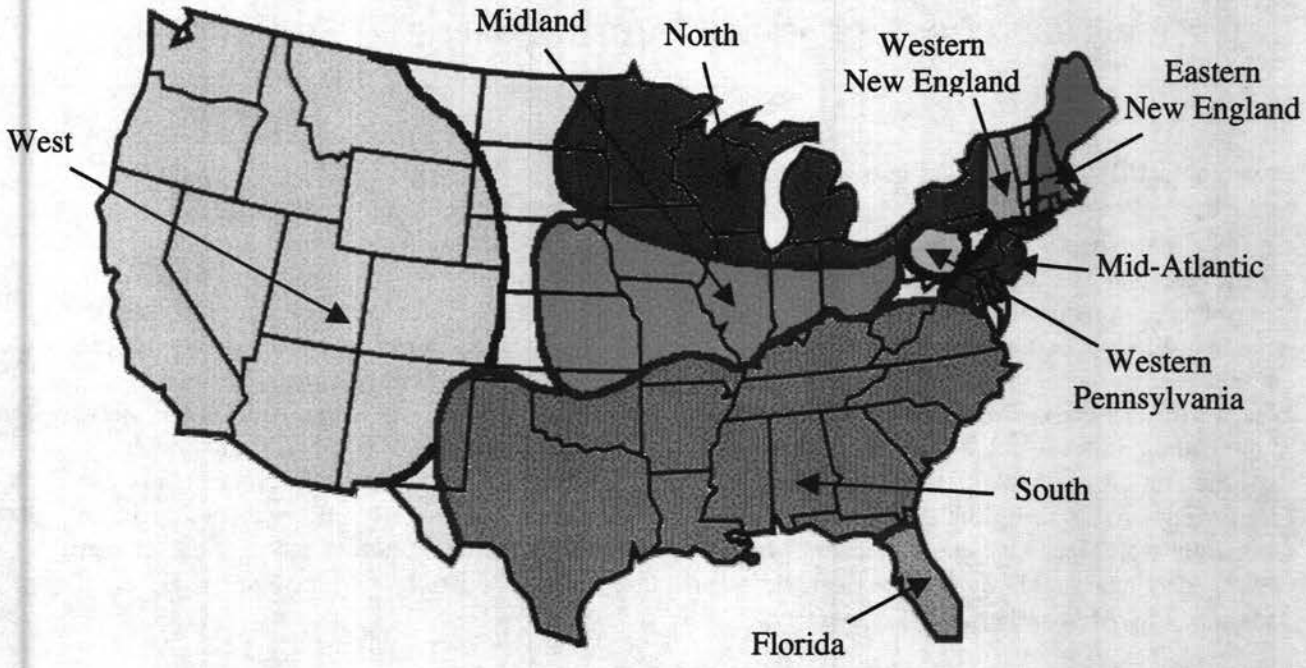


Figure 1. The major dialects of American English, as described by Labov et al. (forthcoming).

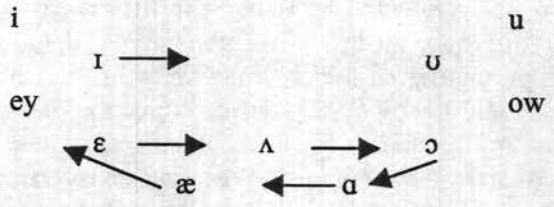


Figure 2. The Northern Cities Chain Shift.

The Southern dialect of American English is characterized by the Southern Vowel Shift (Labov, 1998). The primary feature of this shift is the fronting of the back vowels [u] and [ow]. In addition, the front lax vowels [ɪ] and [ɛ] are raised in Southern American English and the front tense vowels [i] and [ey] are lowered. The Southern Vowel Shift is shown in Figure 3. The Southern dialect is also characterized by monophthongization of the diphthongs [ɑy] and [oy] (Thomas, 2001).

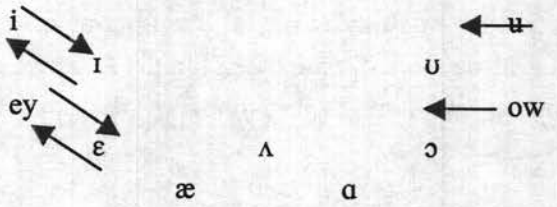


Figure 3. The Southern Vowel Shift.

The common feature of the "third dialect" of American English is the merger of the low-back vowels [ɑ] and [ɔ], creating homophones of such pairs of words as *caught* and *cot* or *Dawn* and *Don*

(Labov, 1998). The subdialects of the "third dialect" also have some unique features of their own. Other features of Eastern New England include raising of the nucleus in the diphthongs [ɑy] and [ɑw] (Thomas, 2001). Western New England, on the other hand, reflects some components of the Northern Cities Chain Shift with some raising of [æ], fronting of [ɑ], and backing of [ɛ] (Boberg, 2001; Thomas, 2001). Western speech is characterized by the low-back merger and by [u] fronting (Labov et al., forthcoming; Thomas, 2001). Unlike Southern back-vowel fronting, however, the Western pattern is typically limited to fronting of [u]. The Midland dialect is the least marked of the regional American English varieties, exhibiting no distinct features other than the Third Dialect [ɑ] ~ [ɔ] merger.

As mentioned above, the Mid-Atlantic dialect does not exhibit the "third dialect" [ɑ] ~ [ɔ] merger and in fact, the two vowels are found to be more distinct due to [ɔ] raising (Labov, 1994; Thomas, 2001). [æ] also exhibits raising in some words, but not others, in the Mid-Atlantic region due to a maintenance of a historical contrast between long and short [æ] (Labov, 1994; Thomas, 2001).

Spoken Language Corpora with Dialect Variation

A number of factors must be considered when designing and collecting a speech corpus, including the demographics of the talkers and the interviewer(s), the recording equipment and conditions, and the types of speech materials to be collected. Table 1 provides examples of each of these factors. Talker demographics include age, gender, socioeconomic status, race, ethnicity, level of education, residential history, and linguistic experience. Residential history can include the region of origin of the talker as well as how many different places he or she has lived. Linguistic experience includes the talkers' native language and any foreign language experience and exposure. The experimenter must decide whether or not to control for each of these variables. The decision to include specific variables or exclude others is related to the ultimate goals of the corpus. For example, if the primary use of the corpus will be comparisons between certain linguistic forms across gender, the experimenter would want to design a corpus that is balanced for gender. If, however, the experimenter is interested only in the speech of female newscasters, the corpus could be limited to female talkers.

Table 1. Some factors to consider in designing a corpus of spoken language.

Factor	Examples
Talker Demographics	age, gender, socioeconomic status, race, ethnicity, level of education, residential history, linguistic experience
Interviewer Demographics	insider vs. outsider
Recording Conditions	fieldwork recordings on tape (analog or digital), telephone recordings, digital recordings in a sound-attenuated booth
Speech Materials	spontaneous speech, interview speech, read speech ("lab speech"), "motherese"

Research in social psychology has shown that talkers often accommodate their speech to that of their interlocutor (Giles & Powesland, 1997; Trudgill, 1998). Therefore, given the potential for stylistic shifts in a talker's speech due to perceived social differences between the interviewer and the talker, the demographics of the interviewer(s) must also be considered. In addition to considering the same factors described above for the talkers, the experimenter must also determine whether the interviewers should be "insiders" or "outsiders" to the community or communities that the speech

corpus represents (Feagin, 2002; Wolfram & Fasold, 1997). As a result of these issues, some corpora rely on a number of different interviewers with different backgrounds, while others rely on only a single interviewer.

The experimenter must also balance two aspects of the recording equipment and conditions: quality of the recordings and speaking style. While sound-attenuated booths and high quality digital recording equipment lead to high quality recordings, such a formal setting also typically results in "lab speech" (Labov, 1972b; Rischel, 1992). On the other hand, fieldwork practices typically result in more natural, conversational speech, but rely on poorer quality recording devices, such as analog audio tape, or have more background noise (Plichta & Mendoza-Denton, 2001). In order to determine the most appropriate recording conditions for a given corpus, the goals of the project must be considered. For example, if the recordings are to be used to document lexical and phonological variation, field recordings may be acceptable. However, if the recordings are to be used in acoustic analyses or playback experiments with naïve listeners, higher quality recordings would be preferred.

Finally, the experimenter must decide what kinds of speech materials to collect. Traditional sociolinguistic research is based on interview speech in which the informants respond to questions to elicit specific lexical items or are asked open-ended questions about childhood games, near-death experiences, or the local community. Most speech perception and spoken word recognition research, on the other hand, is based on read speech produced in the laboratory in order to control for the lexical, segmental, and prosodic content of the utterances. Numerous studies have shown that speaking style (e.g., read speech vs. conversational speech) affects the degree to which certain regional or ethnic dialect variables are produced, with fewer stigmatized forms appearing in read speech than in interview speech (Labov, 1972a). Many recent speech corpora contain some samples of both read and spontaneous speech in order to provide both "natural" and linguistically-controlled utterances.

A number of corpora currently exist that contain variation due to regional and ethnic dialects. A summary of the features of these five corpora is shown in Table 2. Because these corpora were collected with different goals and intended uses, they all have some strengths and some weaknesses related to the factors described above. For example, the Dictionary of American Regional English (DARE) project includes fieldwork recordings of interviews with individuals in more than 1000 communities across the United States collected between 1965 and 1970 (Hall & von Schneidemesser, 2004). The interviews included more than 1800 questions and also included a reading of the Arthur the Rat passage, a short narrative designed to elicit regional phonological variation. The talkers in the DARE interviews differed in terms of their age, ethnicity, socioeconomic status, and gender. The recordings have been used primarily in the production of the Dictionary of American Regional English (1985-) which describes lexical variation in the United States in detail in five volumes. The strengths of this corpus include the large number of talkers and the large amount of speech from each talker. The weaknesses include its poor recording quality and its uneven distribution of talker demographics.

Another example of a speech corpus that includes traditional sociolinguistic interviews is the recent SLX Corpus of Classic Sociolinguistic Interviews (Strassel, Conn, Wagner, Cieri, Labov, & Maeda, 2003). This corpus includes eight interviews with a total of nine different talkers producing a range of utterances including narratives, interview responses, and word lists. The primary strength of the SLX corpus is its utility as a pedagogical tool for training sociolinguistic fieldworkers. As with the DARE interviews, the recording quality of the materials is relatively poor, although the current corpus contains digital files recorded from the original fieldwork tapes. Another weakness of the SLX corpus is the relatively small number of talkers included.

The Santa Barbara Corpus of Spoken American English (DuBois, Chafe, Meyer, & Thompson, 2000) is another source for speech samples containing regional and ethnic variation. The

Santa Barbara corpus includes hundreds of recordings of "natural" speech including conversations, political speeches, classroom lectures, and bedtime stories. The talkers differ in terms of their age, gender, ethnicity, region of origin, and socioeconomic status. The materials from the Santa Barbara corpus are particularly well-suited for studies of prosodic variation and discourse analysis. The main strengths of the corpus are its wide range of speaking styles and speech materials. Its weaknesses include variable recording conditions and an uneven distribution of talker demographics.

One set of corpora that explicitly matched talkers for their regional dialect is the CallFriend project (Canavan & Zepperlen, 1996a, b). One of these corpora includes 60 telephone conversations between two speakers of Southern American English and the other includes 60 telephone conversations between two speakers of non-Southern varieties of American English. Together these corpora provide an excellent source for materials for spoken word recognition research and, in conjunction with the 13 other CallFriend corpora from the Linguistic Data Consortium in different languages including Arabic, Spanish, and Vietnamese, language identification research. The main strength of this corpus is the large number of talkers included. Its weaknesses include the limited bandwidth of the telephone recordings and the fact that the assignment of talkers to the Southern or non-Southern corpus was based on the speech of each talker, not his or her residential history.

Table 2. Summary of existing speech corpora with regional dialect variation.

Corpus	Description
Dictionary of American Regional English	• Fieldwork interviews in 1000 communities in the United States • 1800-item interview questionnaire and a read passage • Used to create the five volume Dictionary of American Regional English
SLX Corpus of Classic Sociolinguistic Interviews	• 8 sociolinguistic interviews with a total of 9 talkers • Narratives, interview responses, word lists • Used as a pedagogical tool for training sociolinguistic fieldworkers
Santa Barbara Corpus of Spoken American English	• Hundreds of recordings of "natural speech" • Settings include conversations, political speeches, classroom lectures, and bedtime stories • Used for studies of prosodic variation and discourse analysis
CallFriend Corpora	• 60 telephone conversations between two American non-Southerners • 60 telephone conversations between two American Southerners • Used for spoken word recognition and automatic language identification
TIMIT Acoustic-Phonetic Continuous Speech Corpus	• 630 talkers ranging in age, ethnicity, gender, education, and regional background • Read sentence materials • Used for spoken word recognition, acoustic analysis of regional variation, perceptual dialect categorization, and automatic dialect identification

The only existing corpus of regional variation in the United States that obtained high quality recordings from a set of talkers in a sound-attenuated booth is the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986; Zue, Seneff, & Glass, 1990). The TIMIT corpus contains recordings of 630 talkers who each read 10 different sentences. Age, gender, ethnicity, level of education, height, and regional dialect are provided for each talker. The TIMIT corpus was originally designed for use in automatic speech recognition research although it has also been used recently in acoustic analyses of regional and gender-based variation (Byrd, 1994; Clopper & Pisoni, 2004b), perceptual dialect categorization (Clopper, Conrey, & Pisoni, in press; Clopper & Pisoni, 2004a, b), and automatic dialect classification (Rojas, 2002). The strengths of the TIMIT corpus include the high quality of the recordings and the large number of talkers. The main weakness of the TIMIT is the limited amount of speech from each talker. In addition, the regional labels assigned to the talkers do not accurately reflect the major regional varieties of American English that Labov et al. (forthcoming) have proposed and it is unclear what criteria were used to assign each talker his or her regional label.

Each of the five corpora described above was designed for a different purpose and the strengths and weaknesses of each corpus reflect those varied goals. A corpus such as the DARE recordings covers a large amount of geographic and lexical territory, but is limited by the quality of the recordings. The TIMIT corpus, on the other hand, provides high quality recordings for a large number of talkers, but the speech materials from each talker are severely limited.

Corpus Design and Collection

The Nationwide Speech Project (NSP) corpus was designed to provide a large amount of speech from male and female talkers representing a number of different regional varieties of American English for use in acoustic analyses and perceptual tasks with naïve listeners. Nearly an hour of speech is available for each talker in a range of speaking styles from isolated read words to interview speech. The read words, sentences, and passages can be used in acoustic analyses and playback experiments when it is desirable to have identical linguistic content across all of the talkers. The interview speech samples can be used for projects in which more "natural" or continuous speech samples are desired. To allow for precise acoustic measurements and reduce the effects of non-linguistic artifacts in the recorded stimulus materials, the recordings were made using high quality digital equipment in a sound-attenuated booth. Finally, the demographic variables of the talkers were strictly controlled such that the resulting corpus includes speech from a relatively homogeneous population of talkers that vary only by gender and region of origin.

Stimulus Materials

Four different kinds of speech materials were collected from each talker in the NSP corpus: isolated words, isolated sentences, passages, and interview speech. Table 3 shows examples of the materials collected for the NSP corpus. The entire list of materials is provided in Appendix A. The isolated words were divided into three materials sets: hVd words, CVC words, and multisyllabic words. The hVd words consisted of five repetitions of each of 10 American English vowels in the hVd context: *heed*, *hid*, *hayed*, *head*, *had*, *hod*, *hud*, *hoed*, *hood*, and *who'd*. The CVC wordlist was composed of 76 monosyllabic English words. Each of the 14 monophthongal and diphthongal vowels in American English was included at least four times in the CVC list and the following consonantal context for each vowel was varied to include liquids, nasals, and voiceless and voiced obstruents. The multisyllabic word list was a subset of 112 of the stimulus materials originally designed by Carter and Clopper (2002) for their study of word reduction by normal-hearing adults. The words in the list were

balanced for number of syllables (two, three, or four), location of primary stress (first, second, or third syllable), and morphological complexity (monomorphemic or polymorphemic). All of the words in the CVC and multisyllabic lists were highly familiar and received a familiarity rating of at least 6.0 (on a seven-point scale) by Indiana University undergraduates (Nusbaum, Pisoni, & Davis, 1994).

Table 3. The speech materials collected from each talker in the NSP corpus.

	Materials Set	**Number of Tokens**	**Examples**
Words	hVd Words	10	heed, hid, head
	CVC Words	76	mice, dome, bait
	Multisyllabic Words	112	alfalfa, nectarine
Sentences	High Probability Sentences	102	Ruth had a necklace of glass beads. The swimmer dove into the pool.
	Low Probability Sentences	52	Tom has been discussing the beads. She might consider the pool.
	Anomalous Sentences	52	Bill knew a can of maple beads. The jar swept up the pool.
Passages	Rainbow Passage	1	When sunlight strikes the raindrops in the air …
	Goldilocks Passage	1	Once upon a time, there were three bears …
Interview Speech	Interview Speech	(5 minutes)	hometown, travel experiences
	Targeted Interview Speech	10	sleep, shoes, math

The read sentence materials were also divided into three materials sets: high probability sentences, low probability sentences, and semantically anomalous sentences. The high probability and low probability sentences were taken from the Speech Perception in Noise (SPIN) test (Kalikow, Stevens, & Elliott, 1977). Examples of the high probability and low probability sentences are shown in Table 3. The SPIN sentences range in length from five to eight words and are phonetically balanced with respect to phoneme frequency in English. High probability sentences are defined as having a final target word that is predictable from the preceding semantic content of the sentence. Low probability sentences have a final target word that is not predictable from the preceding sentence context. The low probability sentences in the SPIN test were created by placing each high probability target word at the end of one of several generic sentence contexts such as, "I did not know about the _____." Thus, in the original SPIN test, all of the high probability sentences were paired with a low probability sentence with the same target word. For the NSP corpus, 102 high probability sentences and 52 low probability sentences were selected from the SPIN test. The low probability sentences were each paired with a high probability sentence with the same target word. The remaining 50 high probability sentences were not paired with a low probability sentence.

The anomalous sentence list was created specifically for the NSP corpus. Using the high probability sentences as a syntactic frame, each content word was replaced with a different content word of the same syntactic class (e.g., noun, verb, or adjective). The target word in each sentence was left unchanged. The resulting utterances were semantically anomalous but syntactically correct sentences. Examples of the anomalous sentences are shown in Table 3. Each anomalous sentence was structurally parallel to one of the high probability sentences. In addition, the target words of the 52 anomalous sentences were matched with the target words of the 52 low probability sentences and the

corresponding 52 high probability sentences. In order to ensure that the anomalous sentences were roughly equivalent in their semantic anomaly, all of the sentences were presented visually to a group of naïve participants who were asked to rate them on a seven-point sensible/strangeness scale. Sentences rated more than one standard deviation above or below the mean were revised (see Clopper, Carter, Dillon, Hernandez, Pisoni, Clarke, Harnsberger, & Herman, 2001).

In addition to the three word lists and the three sentence lists, each talker also read two passages: the first paragraph of the Rainbow Passage (Fairbanks, 1940) and the entire Goldilocks passage (Stockwell, 2002). The Rainbow Passage has been used in a variety of acoustic and perceptual studies of speech, including investigations of talker differences (e.g., Gelfer & Schofield, 2000) and the speech of clinical populations (e.g., Baker, Ramig, Johnson, & Freed, 1997; Hillenbrand & Houde, 1996; McHenry, 1999; Sapienza, Walton, & Murry, 1999). The Goldilocks passage was written to include words and features that would be likely to reveal dialect variation. This passage has been used in the United Kingdom in sociolinguistic studies of language variation and attitudes (Stockwell, 2002). Finally, each talker was recorded while engaged in two conversations with the experimenter. One of the conversations was five minutes in length and included questions about the talker's hometown, extracurricular activities, and travel experiences. The other conversation varied from seven to 12 minutes in length and was designed to elicit certain target words from the talker in relatively natural, conversational speech. Through a series of questions related to specific topics, 10 target monosyllabic words, each containing a different vowel, were elicited from each talker.[3]

Talkers

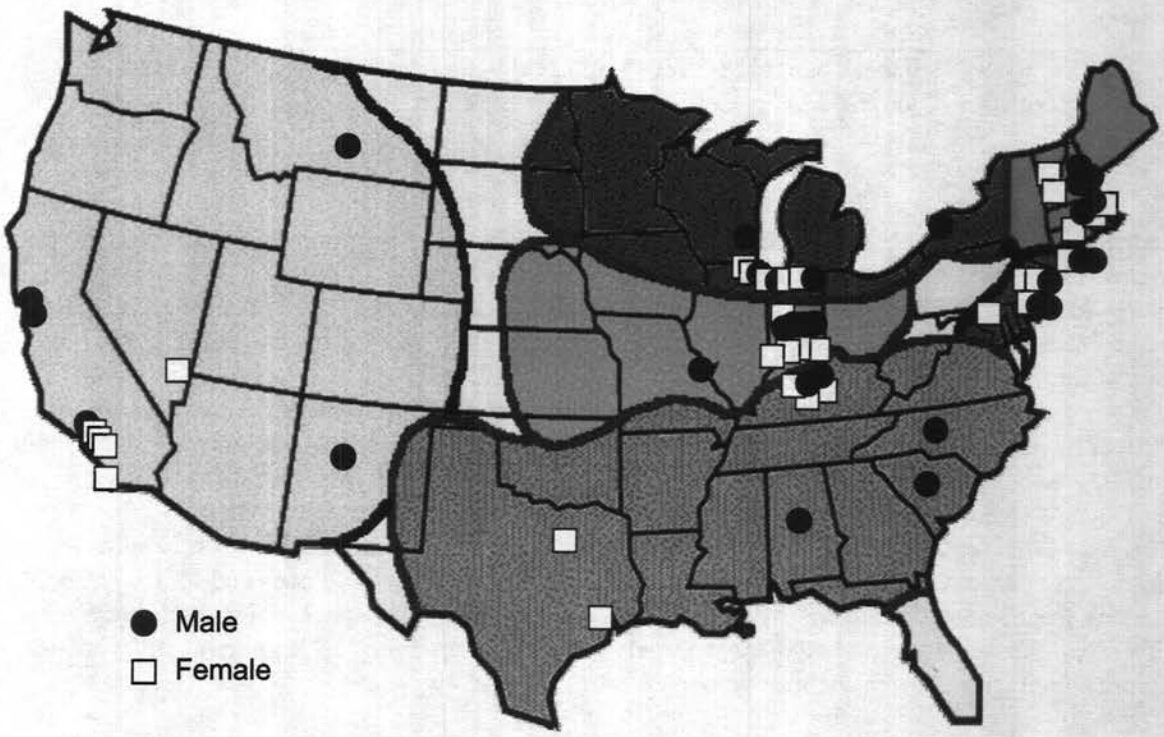


Figure 4. Map of the 60 talkers included in the Nationwide Speech Project corpus. Dark dots represent male talkers and light squares represent female talkers.

[3] I would like to thank Nancy Niedzielski for suggesting this targeted interview task as a means of eliciting specific words in relatively natural continuous speech.

Sixty talkers between the ages of 18 and 25 were recruited from the Indiana University community for participation in the NSP corpus. All of the talkers were monolingual native speakers of American English with no history of hearing or speech disorders. Both parents of each talker were also native English speakers. The 60 talkers included five males and five females from each of six dialect regions of the United States: New England, Mid-Atlantic, North, Midland, South, and West. These six regions were selected based on Labov et al.'s (forthcoming) dialect categories (see Figure 1). In order to qualify for participation, a talker must have lived in a single dialect region for his or her entire life and both of his or her parents must have been raised in that same dialect region. In order to reduce the effects of dialect leveling, each talker had lived in Bloomington, Indiana for less than two years. The map in Figure 4 shows the hometowns for each of the 60 NSP talkers. Male talkers are represented by dark dots and female talkers are represented by light squares.

Procedures

Participants were recorded one at a time by the experimenter (CGC) in a sound-attenuated chamber (IAC Audiometric Testing Room, Model 402). Both the experimenter and the participant sat in the sound booth during testing. During the recording session, the participant was seated in front of a ViewSonic LCD flatscreen monitor (ViewPanel VG151) which mirrored the screen of a Macintosh Powerbook G3 laptop. The participant wore a Shure head-mounted microphone (SM10A) that was positioned approximately one inch from the left corner of the talker's mouth. The microphone output was fed to an Applied Research Technology microphone tube pre-amplifier. The output gain on the pre-amplifier was adjusted by the experimenter while the participant read the Grandfather Passage (Darley, Aronson, & Brown, 1975) as a warm-up before recording began. The output of the microphone pre-amplifier was connected to a Roland UA-30 USB Audio Interface which digitized the signal and transmitted it via USB ports to the laptop where each utterance was recorded in an individual .aiff digital 16-bit sound file at a sampling rate of 44.1kHz. The experimenter held the laptop on her lap and wore headphones connected to the Roland device so that she could hear the same audio signal that was being input to the laptop.

The presentation of the stimulus materials was controlled by the Macintosh laptop using homegrown software. Stimulus items were presented one at a time in 24-point green Courier font on a black background on both the laptop on the experimenter's lap and the LCD screen in front of the participant. The stimulus materials were presented in blocks, such that each participant read all of the CVC words in one experimental block, all of the high probability sentences in another block, and so on, for a total of 10 experimental blocks. Prior to the beginning of each block of trials, the participant was given written instructions on the LCD screen and verbal instructions by the experimenter. Participants were permitted to take breaks between blocks as needed. Each participant within a given dialect received the stimulus materials in a different, random order.

The stimulus items were presented one at a time in random order on the laptop and LCD screens. The durations of the recording intervals varied with each stimulus type and are shown in Table 4. If the participant misread an item or if there was any background noise while the participant read the item, the item was recycled and the trial was presented again at the end of the experimental block.

The entire recording session lasted approximately one hour. The participants received $15 in payment and a Speech Research Laboratory t-shirt for their service.

Table 4. Experiment specifications for the NSP corpus.

Materials Set	Recording Time (sec)	Inter-trial Interval (sec)
hVd Words	2	0.5
CVC Words	2.25	0.5
Multisyllabic Words	3.5	0.5
High Probability Sentences	5	0.5
Low Probability Sentences	5	0.5
Anomalous Sentences	6	0.5
Rainbow Passage	untimed	none
Goldilocks Passage	untimed	none
Interview Speech	300 (5 minutes)	none
Targeted Interview Speech	untimed	none

Acoustic Analysis

In order to confirm that the talkers in the NSP corpus exhibited dialect differences in their speech, an acoustic analysis of the vowels of a subset of the 60 talkers was conducted.

Talkers

The corpus was reduced from 60 to 48 talkers in order to increase the number of trials in the perceptual dialect categorization task (see Chapter 3). Four males and four females from each of the six dialect regions were therefore selected for use in the acoustic analysis and the perceptual experiments described in Chapters 3 and 4. The four talkers of each gender from each dialect region were selected randomly, not based on dialect-related features of their speech.

Stimulus Materials

Table 5. Vowel tokens for acoustic analysis of the Nationwide Speech Project corpus. Tokens followed by (5) were taken from the hVd portion of the corpus, for which five repetitions of each token were available. Tokens followed by (3) were taken from the sentence portion of the corpus, for which three repetitions of each token were available in sentence-final position. Tokens followed by (1) were taken from the CVC portion of the corpus, for which only a single repetition of each token was available.

Vowel	Tokens	Vowel	Tokens
i	heed (5)	ow	hoed (5)
ɪ	hid (5)	ɔ	frogs (3), logs (3)
ey	hayed (5)	a	hod (5)
ɛ	head (5)	ʌ	hud (5)
æ	had (5)	ɑy	guide (1), tide (3)
u	who'd (5)	ɑw	town (1), gown (3)
ʊ	hood (5)	oy	void (1), voice (1), boil (1), coin (1), loyal (1)

A subset of the materials collected from each talker in the NSP corpus were used in an acoustic analysis of the 14 monophthongal and diphthongal vowels of American English: [i, ɪ, ey, ɛ, æ, ɑ, ɔ, ʌ, ow, ʊ, u, ɑy, ɑw, oy] as shown in Table 5. All 50 tokens from the hVd word list were used, for a total of five repetitions of each of 10 vowels. For the remaining four vowels, a combination of

CVC words and sentence-final words from all three sentence lists were used. The selected stimulus items were identical for each talker and the phonetic context was controlled so that nearly all of the vowel tokens occurred before a voiced coronal consonant. The resulting corpus included 69 vowel tokens per talker with 4-6 tokens per vowel.

Procedures

Five acoustic measures were taken from each of the 69 vowel tokens from each of the 48 talkers: vowel duration, first and second formant frequencies at the 1/3 temporal point in the vowel, and first and second formant frequencies at the 2/3 temporal point in the vowel, for a total of 16,560 measurements. All of the measurements were made using the Speech Analysis tool in WaveSurfer 1.6.2 (Sjölander & Beskow, 2004). The speech analysis tool includes a time-aligned waveform, spectrogram with formant tracks for F1, F2, F3, and F4, and an f0 trace. The automatic formant-tracking procedure is computed using a 12-order LPC analysis over a 49ms window with a 10ms frame interval (Sjölander & Beskow, 2004).

For each token, the duration measurements were made first. The onset of the vowel was marked by the onset of voicing for those vowels preceded by a voiceless consonant and by a sudden change in intensity or formant frequency for those vowels preceded by a voiced consonant. The offset of the vowel was marked by the offset of voicing or a sudden drop in intensity, indicating closure. Particularly for those vowels adjacent to a liquid or nasal consonant, vowel onsets and offsets were determined by visual inspection of the waveform and spectrogram as well as by ear. Vowel duration was calculated as the difference between offset and onset of the vowel in milliseconds.

Locations for the formant frequency measures were made based on the duration measures. The first-third temporal point was calculated as one-third of the duration plus the timestamp of the onset. The second-third temporal point was calculated as two-thirds of the duration plus the timestamp of the onset. First and second formant measures were then obtained from the automatic formant tracks at those two exact temporal locations. Formant measures were hand-corrected by the author using the cursor tool as necessary.

A total of 28 vowel tokens (0.8%) were excluded because the talker misread the word (25 tokens) or a recording error occurred (3 tokens). All of the excluded tokens were from the hVd materials set in which trials with disfluencies or mispronunciations were not repeated at the end of the experimental block.

The measurements were hand-checked for outliers prior to any further analysis. A total of 52 measurements out of the total 16,560 (0.3%) were re-checked as potential outliers. Four were errors by the author in measurement, five were typographic errors by the author in data recording, and 22 were formant tracking errors not hand-corrected at the time of the original measurement. These measurements were all corrected prior to the analysis of the data. The remaining 21 potential outliers were found to be due to natural variation in the corpus and these data points were not altered.

A small subset (1%) of the tokens was also remeasured by the author in order to assess reliability. The reliability subset included tokens from all six dialects and all 14 vowels. Correlations between the original measurements and the reliability measurements were highly significant, with r = .993 for the duration measurements and r = .996 for the formant measurements. The reliability analysis and the outlier check suggest that the measurements were highly reliable and accurate.

Results

A summary of the mean formant frequencies for the 11 American English monophthongs [i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, u] for each dialect are shown in Figure 5. The means are based on the formant frequency measures taken at the first-third temporal point. Figure 5 clearly shows the Northern Cities Chain Shift for the Northern talkers, with the fronting and lowering of [ɑ] in *hod* and the raising and fronting of [æ] in *had* relative to the other dialects. In addition, the Southern talkers show fronting of [u] in *who'd* and [o] in *hoed*. The Mid-Atlantic, Midland, and Western talkers also show some fronting of [u] in *who'd*. Finally, the New England, Midland, and Western talkers appear to have a merger or near-merger of [ɑ] and [ɔ] in *hod* and *frogs/logs*.

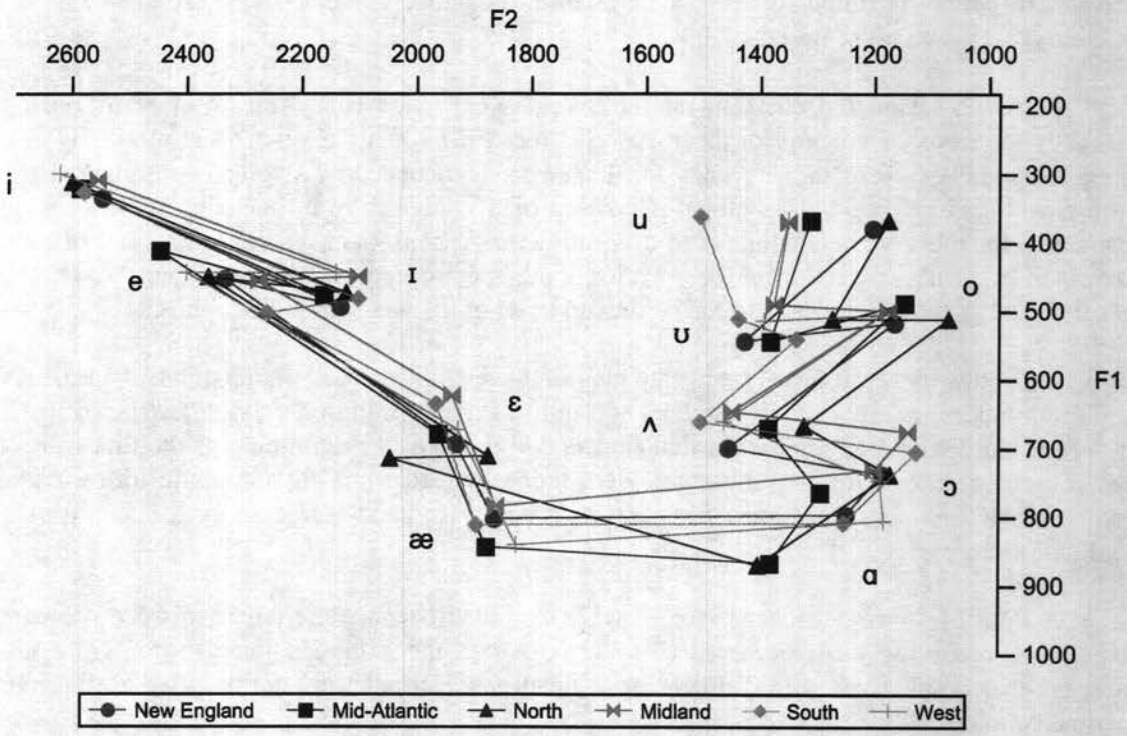


Figure 5. Mean formant frequency values for 11 American English vowels for each of the six dialects. The lines connecting the vowels proceed in a counterclockwise direction as follows: [i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u].

Figures 6 through 11 show all of the tokens for each of the same 11 monophthongal American English vowels for each dialect. In these figures, the tokens have been plotted by gender for clarity. The ellipses in these figures were drawn by hand to include all of the tokens for each vowel. Individual vowel spaces for each talker are shown in Appendix B.

The top panel of Figure 6 shows the data for the New England male talkers. The most striking aspect of this figure is the split between the different talkers in their production of [æ]. While three of the talkers showed raised [æ]s similar to those found in the Northern dialect region, one talker (NE1) retained lowered [æ]s. In addition, the New England males appear to produce a merger or near-merger of [ɑ] and [ɔ].

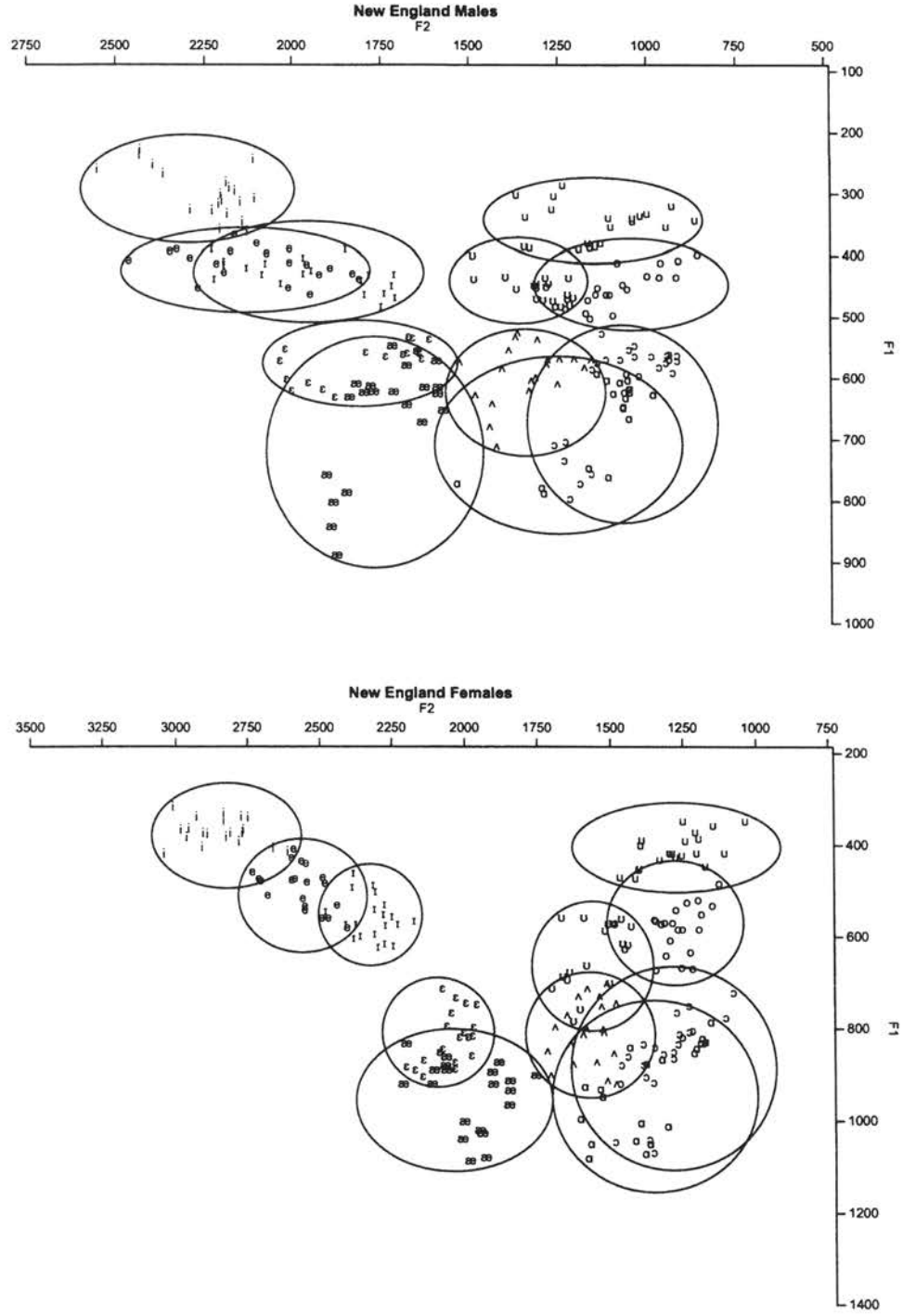


Figure 6. All tokens produced by the New England male (top) and female (bottom) talkers for the 11 vowels [i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u]. The ellipses were hand-drawn to include every token for each vowel.

The data for the New England females are shown in the bottom panel of Figure 6. Like the New England males, the New England females showed two different productions of [æ]. One talker (NE8) had a raised [æ] consistent with the Northern Cities Chain Shift, while the other three talkers maintained a distinction between the nuclei of [æ] and [ɛ]. The New England women also produced a merger or near-merger of [ɑ] and [ɔ], like their male counterparts.

The data for the Mid-Atlantic males are shown in the top panel of Figure 7. The Mid-Atlantic males seem to be the least variable subgroup of talkers in the NSP corpus. In general, the ellipses in Figure 7 are smaller than the ellipses found in the other figures. We also see some evidence for a merger of [ɑ] and [ɔ] among the Mid-Atlantic males. Inspection of the individual vowel spaces in Appendix B suggests that this may be the case for two of the talkers (AT1 and AT3), but that the other two talkers (AT2 and AT5) maintained distinct low-back vowels.

The bottom panel of Figure 7 shows the vowel formant frequency data for the Mid-Atlantic women. Like the Mid-Atlantic males, the Mid-Atlantic females were inconsistent in producing the low-back merger. Only one talker (AT9) showed a merger of [ɑ] and [ɔ], while the other three talkers maintained a distinction between these two vowels. Only one talker (A18) produced the expected raised [ɔ]. The Mid-Atlantic women were also variable in their production of [u]. One talker (A18) produced fronted [u]s, while the other three females produced more backed [u]s. Finally, Figure 7 shows a great deal of overlap between [ʊ] and [ʌ] for the Mid-Atlantic females. An inspection of the individual vowel spaces in Appendix B suggests that this overlap is not due to mergers at the individual talker level, but is due to variation across talkers in the production of these vowels, particularly in F1.

The top panel of Figure 8 shows the data for the Northern males. The Northern Cities Chain Shift appears to be present in all four talkers in this sample. [ɑ] is clearly distinct from [ɔ], due to lowering and fronting of [ɑ]. All four talkers also produced raised and/or fronted [æ]s. In addition, [ɛ] shows some backing to reduce overlap with raised [æ]. An inspection of the individual vowel spaces in Appendix B confirms that the talkers from upstate New York and Wisconsin produced backed [ɛ]s (NO2 and NO4, respectively), while the two talkers from Northern Indiana maintained a more fronted production of [ɛ] (NO3 and NO5). [ʌ] was also backed and shows some overlap with [ɔ].

The data for the Northern females are shown in the bottom panel of Figure 8. The vowels produced by the Northern women also reflect the Northern Cities Chain Shift. Like the Northern men, these women produced lowered and fronted [ɑ]s, raised and fronted [æ]s, and backed [ɛ]s and [ʌ]s. One of the Northern women (NO9) also produced fronted [u]s, while the other three retained backed [u]s.

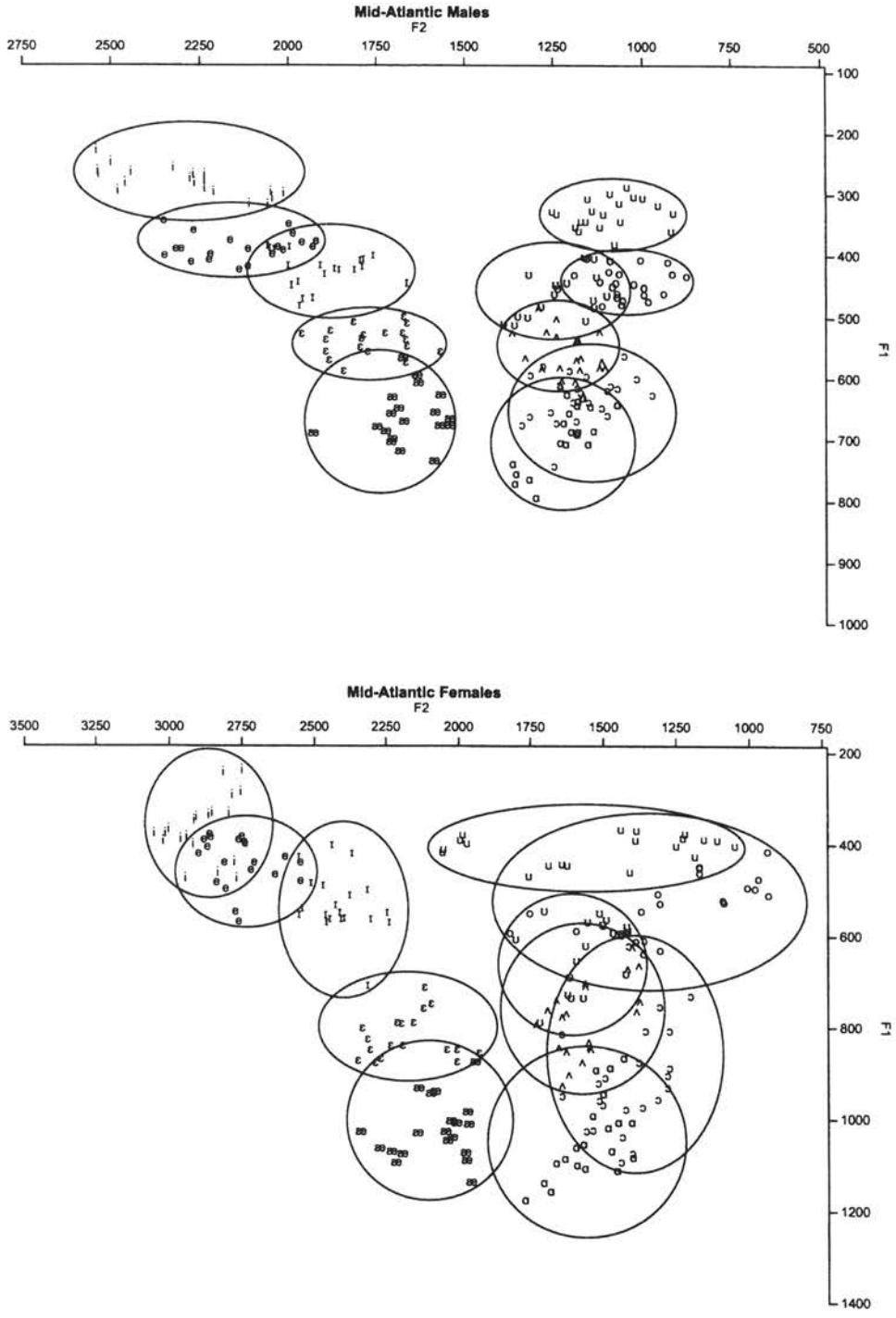


Figure 7. All tokens produced by the Mid-Atlantic male (top) and female (bottom) talkers for the 11 vowels [i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u]. The ellipses were hand-drawn to include every token for each vowel.

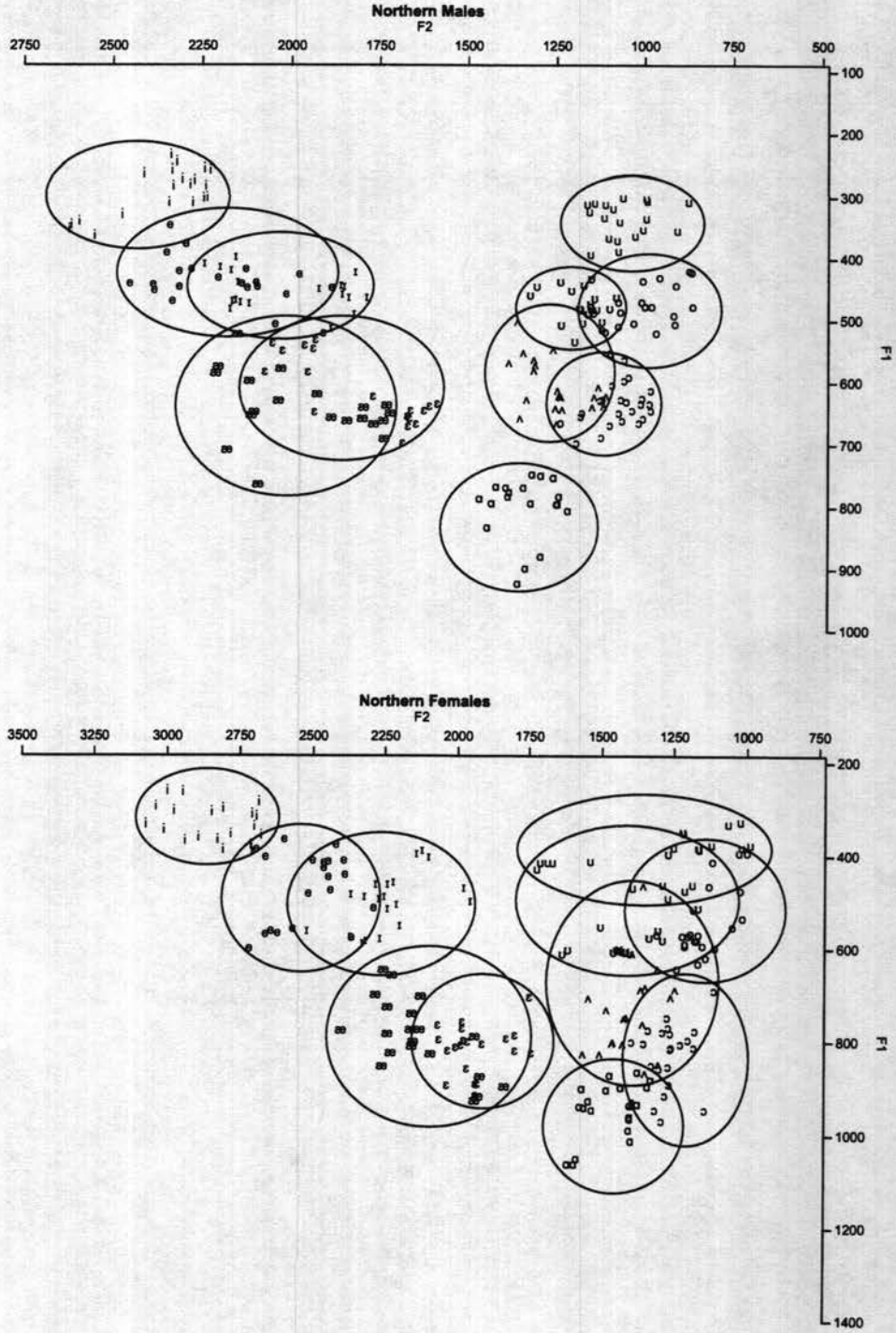


Figure 8. All tokens produced by the Northern male (top) and female (bottom) talkers for the 11 vowels [i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u]. The ellipses were hand-drawn to include every token for each vowel.

The vowel tokens for the Midland male talkers are shown in the top panel of Figure 9. This figure reveals two interesting splits between the talkers. First, two talkers (MI3 and MI4) showed fronted [u]s whereas the other two talkers (M12 and MI1) had more backed [u] productions. In addition, two of the talkers (M12 and MI3) had raised [æ]s, whereas the other two talkers (MI1 and MI4) retained the lower [æ] production. These results are particularly interesting because [u] fronting is associated with the Southern shift while [æ] raising is associated with the Northern Cities Chain Shift. It is somewhat surprising that a single talker (MI3) would exhibit both Southern [u] fronting and Northern [æ] raising. It is also interesting to note in this figure that the vowel [ɪ] is completely encompassed by the vowel [e]. This result is due to an apparent merger of these two vowels (at least in terms of nucleus formant frequency) for MI4.

The vowel tokens for the Midland females are plotted in the bottom panel of Figure 9. Unlike their male counterparts, who individually showed evidence of Southern and Northern features, the Midland women showed very few shifted vowels. An inspection of the individual vowel spaces in Appendix B suggests that the variation in Figure 9 is due to overall differences between talkers and not to individual differences in vowel shifts or mergers, with the exception of one talker (MI6) who produced slightly raised [æ]s. The Midland women as a group also exhibited the merger of [ɑ] and [ɔ] and this merger is evident in the individual vowel spaces for three (MI8, MI9, and MI0) of the four talkers.

Vowel production data for the Southern males are shown in the top panel of Figure 10. [u] and [o] fronting are consistently present in all of the tokens. In the case of [o] fronting, this results in a near-complete overlap of [o] and [ʊ] across talkers. Inspection of the individual vowel spaces in Appendix B, however, suggests that each individual talker maintained a distinction between [o] and [ʊ] in both the nucleus position and in the formant trajectory over the course of the vowel. Like the Midland talkers, the Southern talkers showed highly similar [e]s and [ɪ]s with respect to formant frequencies at the first-third temporal point, but in all four talkers, the trajectories for [e] and [ɪ] were clearly distinct. [e] moved up and front over the course of the vowel, while [ɪ] moved back and down (see Appendix B).

The bottom panel of Figure 10 is a plot of the vowel tokens for the Southern females. The Southern females were by far the most variable subgroup of talkers in the NSP corpus. The overlap across different vowels is quite large for the entire space except for [i]. As a group, the Southern women showed a general trend for [u] and [o] fronting like the Southern men, but individually, the women ranged from virtually no fronting of these vowels by one of the talkers from Kentucky (SO6) to very fronted back vowels by one of the talkers from Texas (SO7).

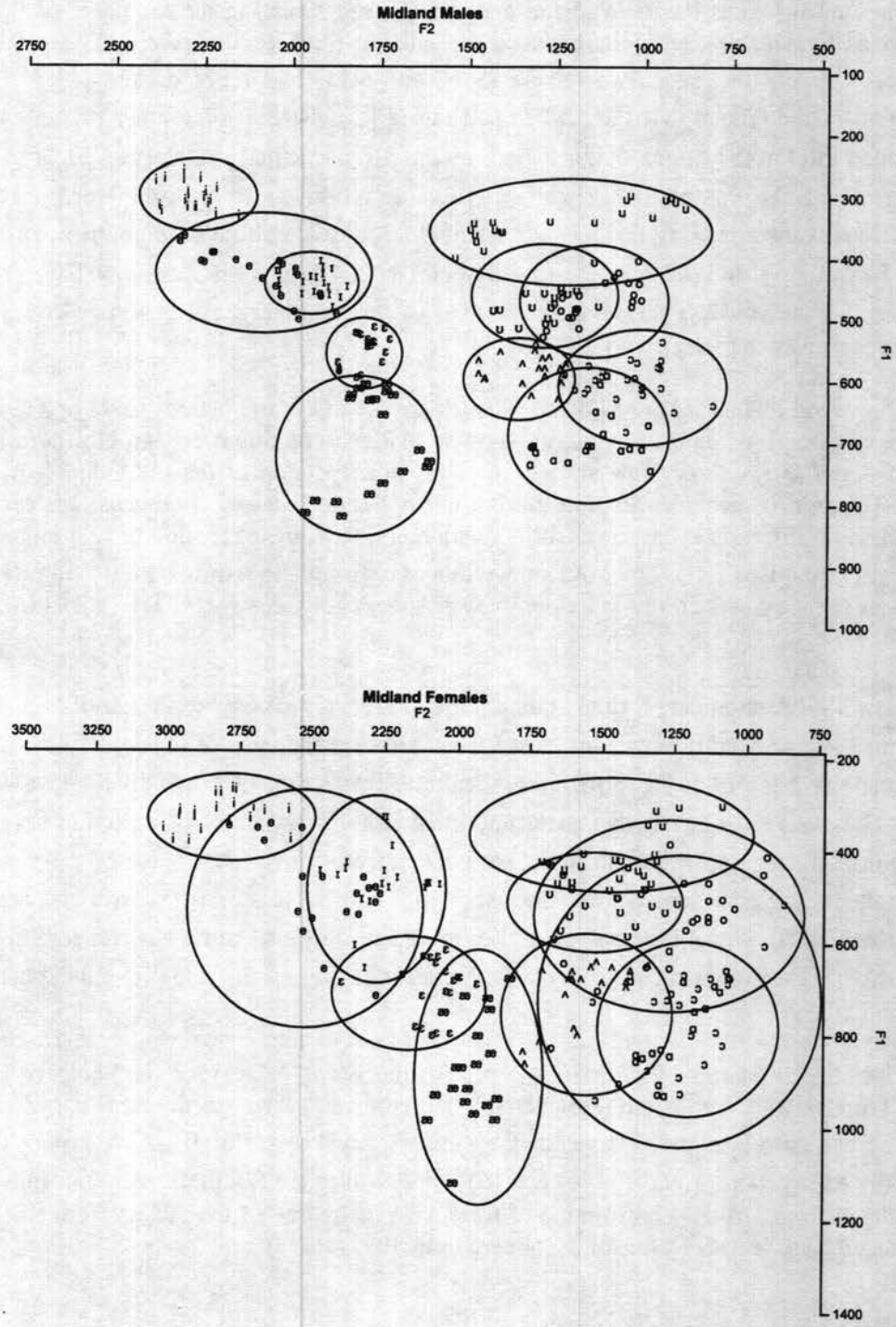


Figure 9. All tokens produced by the Midland male (top) and female (bottom) talkers for the 11 vowels [i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u]. The ellipses were hand-drawn to include every token for each vowel.

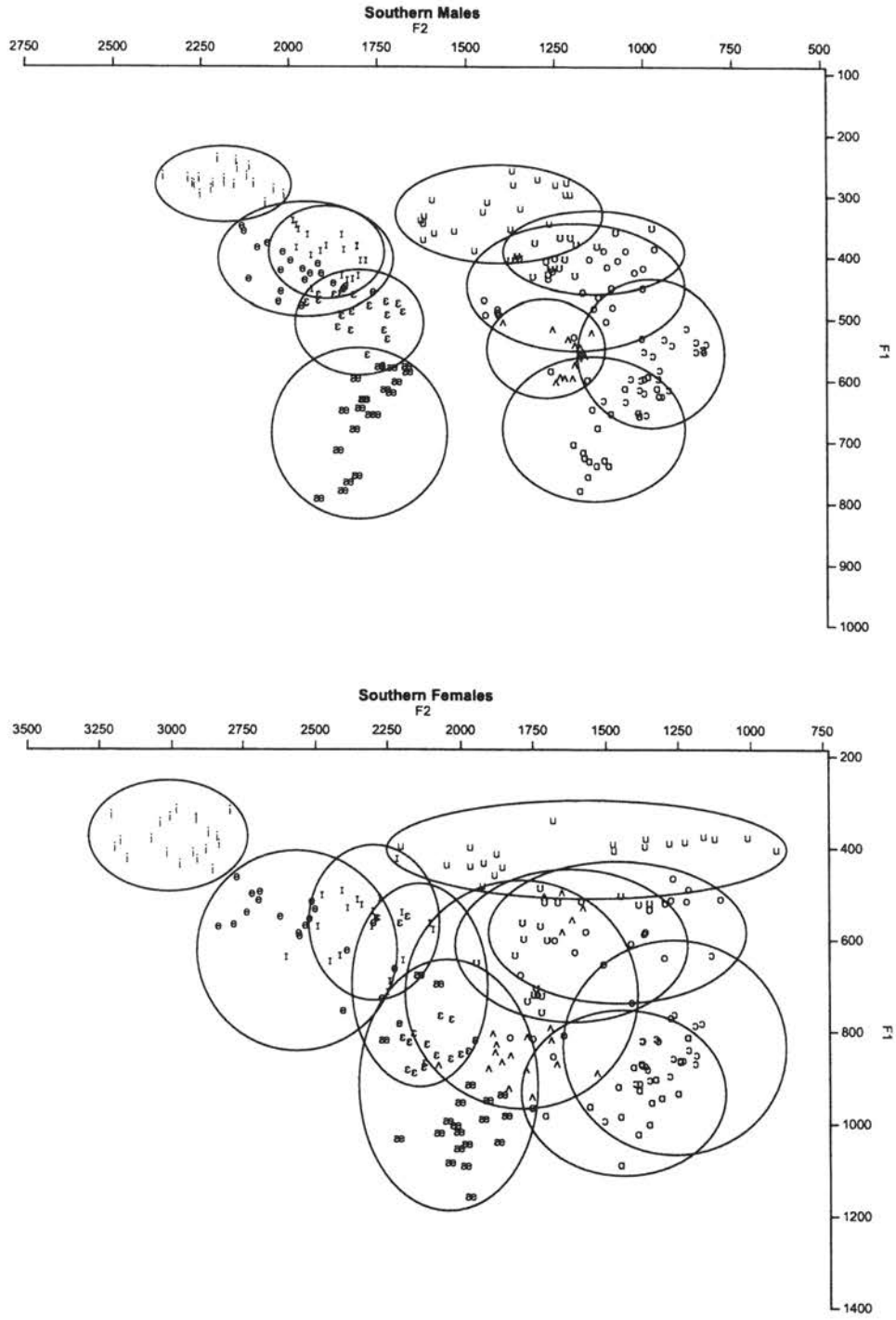


Figure 10. All tokens produced by the Southern male (top) and female (bottom) talkers for the 11 vowels [i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u]. The ellipses were hand-drawn to include every token for each vowel.

The top panel of Figure 11 shows the data for the last group of male talkers, the Westerners. The large variance in F2 for [u], [ʊ], and [o] is due to the talker from Montana (WE4) who produced back vowels with very low second formants. The merger of [ɑ] and [ɔ] is also visible in this figure and

inspection of the individual vowel spaces in Appendix B suggests that all four talkers produced merged or nearly-merged low-back vowels.

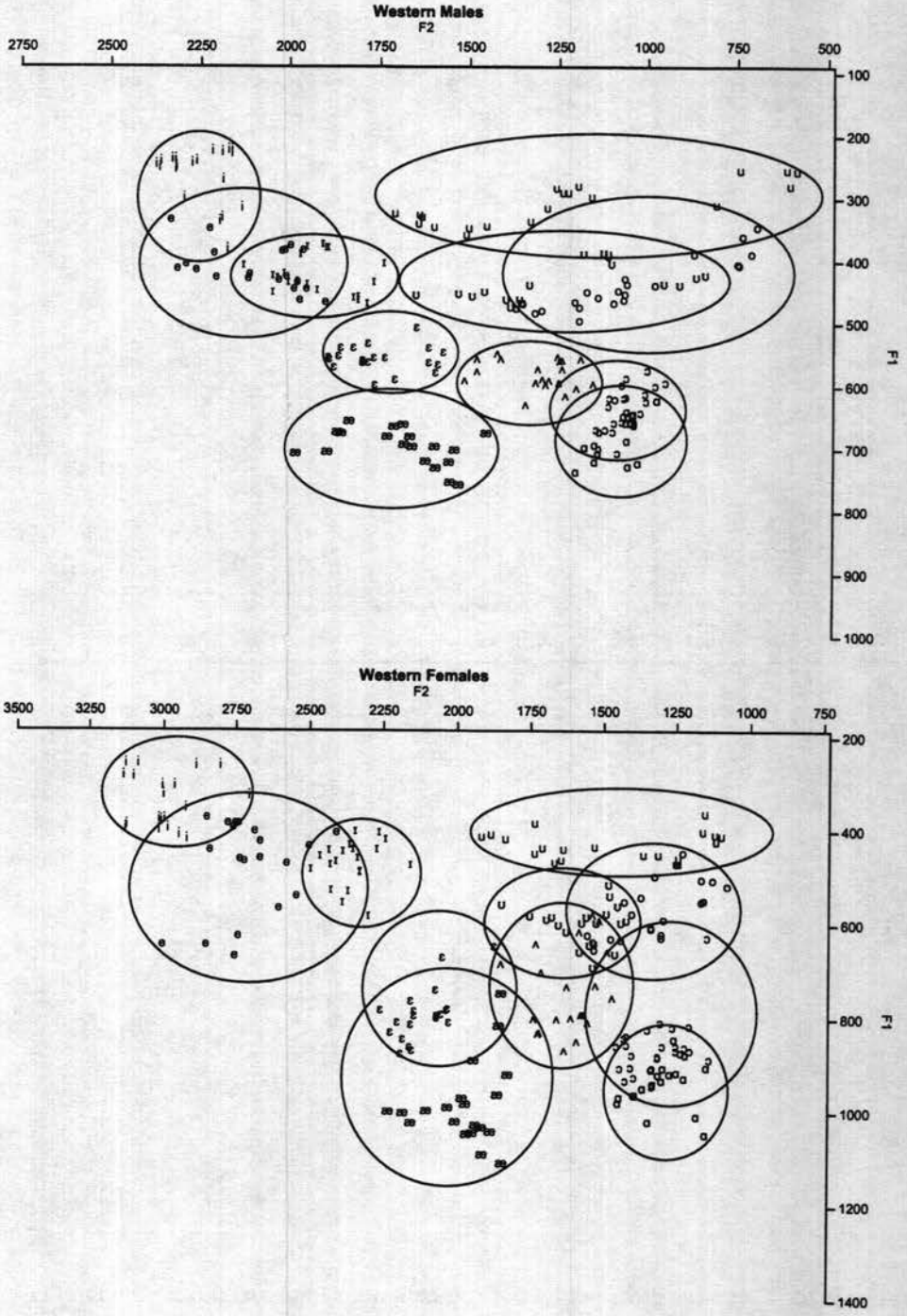


Figure 11. All tokens produced by the Western male (top) and female (bottom) talkers for the 11 vowels [i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u]. The ellipses were hand-drawn to include every token for each vowel.

Finally, the data for the Western female talkers are shown in the bottom panel of Figure 11. Like their male counterparts, the Western women showed a merger of [ɑ] and [ɔ] as a group and individually. In addition, [u] fronting was found in two of the Western females (WE7 and WE9), but not in the others.

In general, the women were more variable within a given dialect than the men. A comparison of the individual vowel spaces in Appendix B, however, suggests that individual women were not any more variable in their productions of specific vowels than individual men, but that across different women within a given dialect, there was simply more variation. Impressionistically, the women in the NSP corpus sounded more heterogeneous than the men in terms of pitch and speaking style, which may explain these differences in the acoustic measures. In addition, the women may have been more different in terms of overall physical size than the men which may have led to greater biological variation in formant frequencies.

Taken together, the formant frequency measures displayed in the figures discussed above and those in Appendix B suggest that the talkers in the NSP corpus did in fact produce some of the characteristic features of their dialect regions described in the sociolinguistic literature (Labov, 1998; Labov et al., forthcoming; Thomas, 2001). In order to quantitatively assess differences in formant frequency due to talker dialect, a series of statistical analyses was conducted.

Prior to the analysis, all of the formant frequency measures were converted to a Bark scale, which reflects the tonotopic map of the human auditory system in which low frequencies have finer resolution than high frequencies (Traunmüller, 1990). Formant frequencies measured in Barks therefore more accurately reflect how speech is perceived, particularly in the second formant range. The conversion equation used is shown in (1), where z indicates frequency in Barks and f indicates frequency in Hertz (Traunmüller, 1990). No corrections were necessary because all of the formant frequencies in the current data set were in the range $2 < z < 20.1$. All of the figures in this section are plotted in Hertz for clarity.

(1) $z = 26.81f / (1960+f) - 0.53$

A repeated measures ANOVA was calculated using vowel category (i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, u, ɑy, ɑw, or oy) as a within-subjects factor and dialect (New England, Mid-Atlantic, North, Midland, South, or West) as a between-subjects factor for each of five measures: vowel duration, F1 (at the first-third temporal point), F2 (at the first-third temporal point), ΔF1 (change in F1 from the first-third to the second-third temporal point), and ΔF2 (change in F2 from the first-third to the second-third temporal point). Because five analyses were computed, the p-value was set to .01 for each ANOVA. The data were collapsed across talker gender for each analysis in order to obtain the most conservative estimate of reliable differences due to talker dialect. Any significant effects in the following analyses would reflect differences between dialects that are robust across variability due to gender. Interactions between dialect and gender are beyond the scope of the present project and will be left for future research.

Duration. The repeated measures ANOVA on vowel duration revealed a significant main effect of vowel ($F(13, 2132) = 174.6$, $p < .001$), a significant main effect of dialect ($F(5, 2132) = 3.2$, $p = .009$), and a significant vowel x dialect interaction ($F(65, 2132) = 2.4$, $p < .001$). The significant main effect of vowel category merely confirms that American English vowels differ in their inherent length and no further analyses on that factor were conducted. The significant main effect of dialect suggests that some dialects have longer or shorter overall vowels than others. A post-hoc Tukey test

on dialect revealed a significant difference between New England and Southern talkers based on vowel duration ($p = .003$). The Southerners had significantly longer vowels than the New Englanders.

The vowel x dialect interaction suggests that the effects of dialect differences on vowel duration are not consistent across all vowels. In order to explore this interaction more closely, a one-way ANOVA on vowel duration with dialect as the factor was computed for each of the 14 vowels. To correct for the large number of analyses, the p-value for the ANOVAs and the post-hoc Tukey tests was set to .001.

Significant main effects of dialect were found for [ɪ] ($F(5, 234) = 7.6, p < .001$), [ɛ] ($F(5, 228) = 8.2, p < .001$), [ʌ] ($F(5, 229) = 9.3, p < .001$), and [ʊ] ($F(5, 229) = 8.3, p < .001$).[4] In all four cases, post-hoc Tukey tests on dialect revealed that the Southerners' vowels were significantly longer than the vowels produced by the New England, Mid-Atlantic, and Western talkers (all $p < .001$). The vowel [ɛ] was also longer for Southerners than Northerners and the vowel [ʌ] was longer for Southern talkers than for Northern and Midland talkers (all $p < .001$). These results suggest that Southerners did not produce generally longer vowels or have an overall slower speaking rate (as indicated by longer vowels), but that the vowel duration differences based on dialect were due to longer lax vowels for Southern talkers than for the other dialect groups.

Nucleus Formant Frequencies. The repeated measures ANOVAs on F1 and F2 measured at the first-third temporal point revealed a significant main effect of vowel ($F(13, 2132) = 1057.8, p < .001$ for F1 and $F(13, 2132) = 1955.9, p < .001$ for F2). These results merely confirm the existence of significant differences in formant frequencies due to vowel category and will not be further analyzed. Both the F1 and the F2 analyses also revealed a significant vowel x dialect interaction ($F(65, 2132) = 4.5, p < .001$ for F1 and $F(65, 2132) = 7.9, p < .001$ for F2). Neither analysis produced a significant main effect of dialect. These findings suggest that the vowel spaces of the different dialects are not globally shifted along either F1 or F2, but that individual vowels are affected differentially by the six dialects. As in the duration analysis above, one-way ANOVAs on F1 and F2 with dialect as the factor were computed for each of the 14 vowels. The p-value was again set at .001 to correct for the large number of post-hoc analyses.

For F1, significant main effects of dialect were found for [e] ($F(5, 234) = 6.0, p < .001$) and [ɑ] ($F(5, 224) = 5.2, p < .001$). [5] Post-hoc Tukey tests revealed that the Northern dialect had significantly lower [ɑ]s than the Midland dialect and that the Southern dialect had significantly lower [e]s than the Mid-Atlantic dialect. Northern [ɑ] lowering is one feature of the Northern Cities Chain Shift and Southern [e] lowering is one feature of the Southern Shift (Labov, 1998).

For F2, significant main effects of dialect were found for [æ] ($F(5, 233) = 7.0, p < .001$), [ɑ] ($F(5, 224) = 12.6, p < .001$), [ɔ] ($F(5, 282) = 6.5, p < .001$), [ʌ] ($F(5, 229) = 4.5, p = .001$), [o] ($F(5, 232) = 7.8, p < .001$), and [u] ($F(5, 234) = 7.0, p < .001$).[6] Post-hoc Tukey tests revealed significant [æ] fronting by the Northern talkers compared to the New England, Midland, and Western talkers,

[4] Increasing the p-value to .01 leads to one additional significant result for the duration analysis. Dialect differences were significant for [ɑw] at $p < .01$.

[5] Increasing the p-value to .01 leads to one additional significant result for the F1 analysis. Dialect differences were significant for [æ] at $p < .01$.

[6] Increasing the p-value to .01 leads to one additional significant result for the F2 analysis. Dialect differences were significant for [ʊ] at $p < .01$.

significant [ɑ] fronting by the Northern talkers compared to the New England, Midland, Southern, and Western talkers, and significant [ʌ] backing by the Northern talkers relative to the Southern talkers (all $p < .001$). These findings are also consistent with the Northern Cities Chain Shift. The Mid-Atlantic talkers also produced fronted [ɑ]s relative to the Midland and Western talkers (both $p < .001$). In addition, post-hoc Tukey tests revealed significant fronting of [o] and [u] for Southern talkers as compared to Northerners (both $p < .001$). Southern talkers also had significantly more [o] fronting than Mid-Atlantic talkers and significantly more [u] fronting than New England talkers (both $p < .001$). These findings are consistent with the Southern Vowel Shift (Labov, 1998). Finally, the Mid-Atlantic talkers were shown by post-hoc Tukey tests to have more fronted [ɔ]s than Midland and Southern talkers (both $p < .001$). Given the reported tendency for [ɔ] raising in the Mid-Atlantic (Labov, 1994; Thomas, 2001), this result was somewhat surprising. However, the lexical items used to obtain this measure of [ɔ] (*frogs* and *logs*) were among a small set of words that have been reported to show large variation between regional varieties of American English and do not always pattern with other words containing the same vowel (Trager, 1930; Wells, 1982). Thus, this effect may be lexically-specific. In any case, we found that the Mid-Atlantic fronted [ɔ] was still quite distinct from the Mid-Atlantic [ɑ], as predicted.

A summary of these results is shown in Figure 12 which contains a plot of the mean values for each dialect for each of the vowels [e, æ, ɑ, ɔ, o, u]. The significant differences revealed by the Tukey tests are indicated by the dashed square brackets. Note that in this figure, the Midland [ɑ] is located on top of the New England, North, and West [ɔ]s and that the Mid-Atlantic [ɔ] is quite close to the New England and South [ɑ]s.

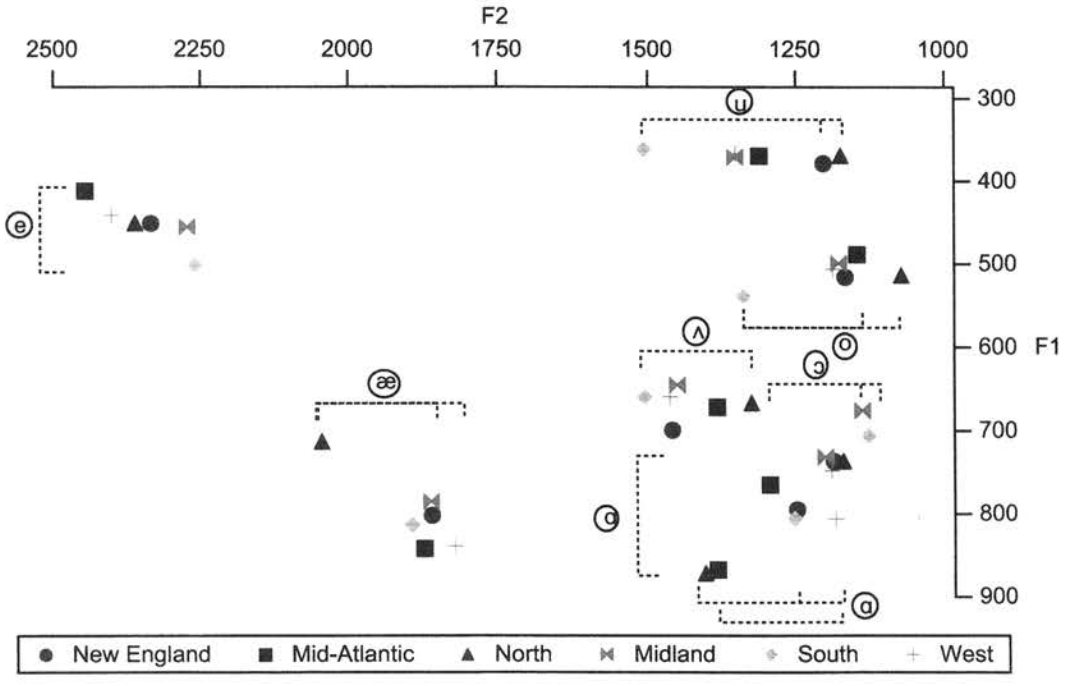


Figure 12. Mean formant frequency values for each dialect for the vowels [e, æ, ɑ, ɔ, ʌ, o, u]. Significant differences based on Tukey tests at the .001 level are depicted by the dashed square brackets.

Formant Trajectories. Repeated measures ANOVAs on ΔF1 and ΔF2 revealed significant main effects of vowel ($F(13, 2132) = 56.6$, $p < .001$ for ΔF1 and $F(13, 2132) = 740.3$, $p < .001$ for ΔF2). This finding merely confirms that different vowel categories exhibit different degrees of diphthongization and will not be analyzed further. Both ANOVAs also revealed a significant vowel x dialect interaction ($F(65, 2132) = 2.9$, $p < .001$ for ΔF1 and $F(65, 2132) = 3.7$, $p < .001$ for ΔF2). The main effect of dialect was not significant in either the ΔF1 or the ΔF2 analysis. These results suggest that the dialects of American English examined here do not differ in their overall diphthongization, but that the effects of dialect on formant trajectory are different for different vowels. To assess the interactions in more detail, one-way ANOVAs on ΔF1 and ΔF2 with dialect as a factor were conducted for each of the 14 vowels. Due to the large number of calculations, the p-value for this set of analyses was set to .001.

For ΔF1, a main effect of dialect was found for [e] ($F(5, 234) = 5.4$, $p < .001$), [ɛ] ($F(5, 228) = 7.7$, $p < .001$), [æ] ($F(5, 233) = 21.2$, $p < .001$), and [ʊ] ($F(5, 229) = 6.7$, $p < .001$). Post-hoc Tukey tests revealed significantly more movement in the F1 of [e] for the Southern talkers than the Mid-Atlantic talkers ($p < .001$). Figure 13 shows the mean trajectories of each dialect for [e] from the first-third temporal point to the second-third temporal point. As is shown in Figure 13, the difference in ΔF1 between the Southern and Mid-Atlantic talkers reflects a compensation for the lower nucleus position of the Southern vowel, as described above; at the two-thirds temporal point in [e], all of the talkers have achieved a similar F1 target.

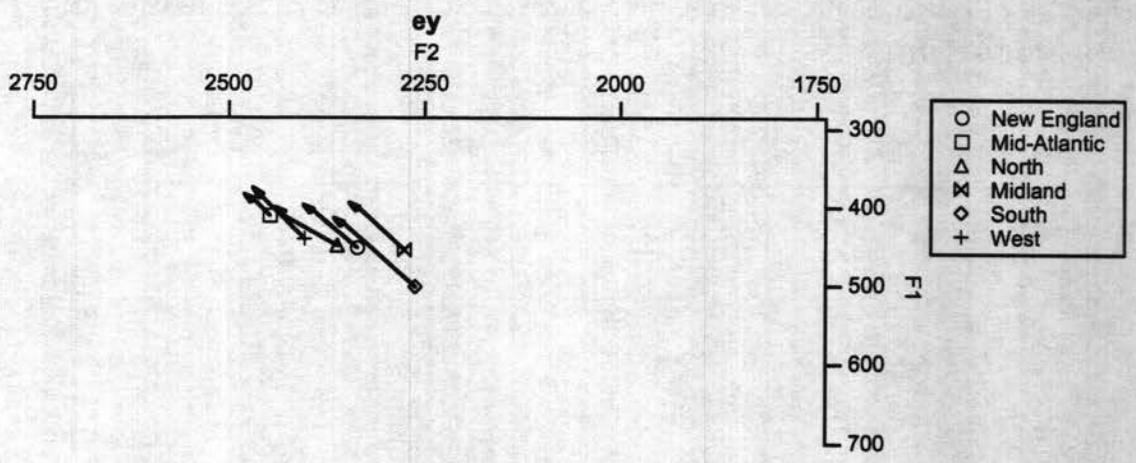


Figure 13. Mean formant frequency trajectories for each dialect for the vowel [e].

Similarly, for [æ], post-hoc Tukey tests revealed significant differences in ΔF1 between the Northern talkers and all of the other dialect groups (all $p < .001$). Recall that one component of the Northern Cities Chain Shift is the raising and fronting of [æ]. As shown in Figure 14, the mean trajectory of [æ] for the Northern talkers is back and down, while the mean trajectory of [æ] for the other dialects and back and up. The significant difference in trajectory between the Northerners and the other talkers is due to a similar offglide target for all six dialects than must be attained through lowering for the Northerners and raising for the other five dialects.

Post-hoc Tukey tests on the F1 trajectory of [ɛ] revealed significantly less F1 movement for Southerners than for Northerners or Westerners (both $p < .001$). As shown by the mean formant

trajectories for each dialect for the vowel [ɛ] in Figure 15, this significant difference is due to a difference in direction of the trajectory for the Southerners. The Southern [ɛ] moves mostly back, whereas the Northern and Western [ɛ]s move mostly up.

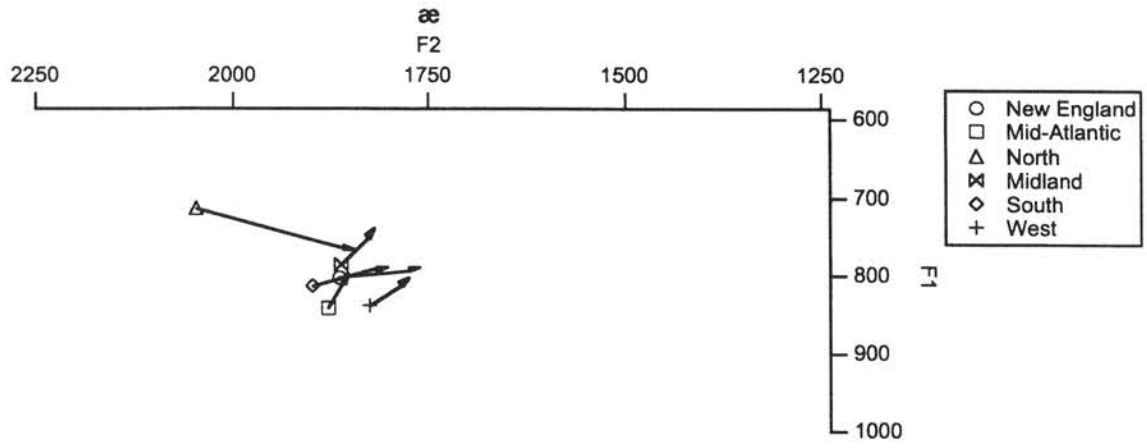


Figure 14. Mean formant frequency trajectories for each dialect for the vowel [æ].

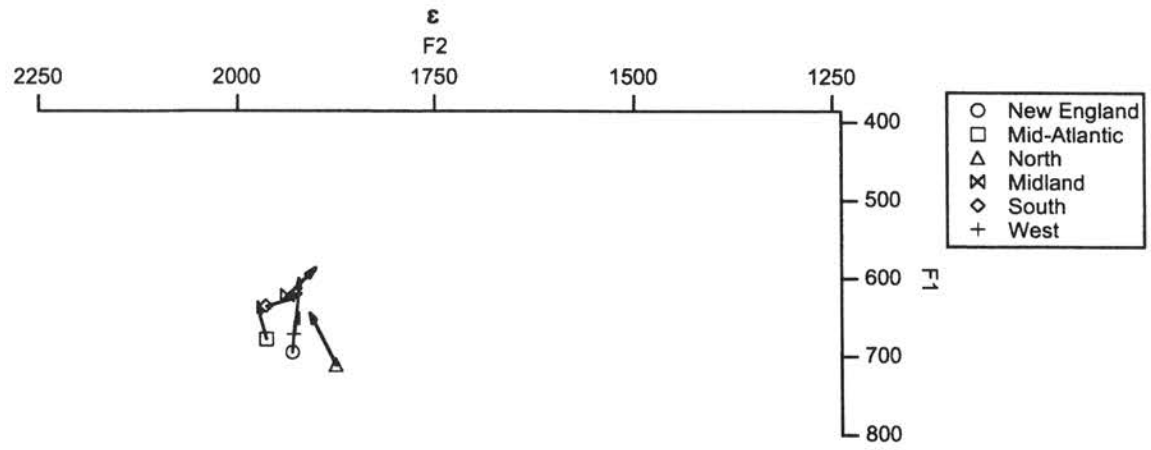


Figure 15. Mean formant frequency trajectories for each dialect for the vowel [ɛ].

Finally, for [ʊ], post-hoc Tukey tests revealed significant differences in F1 trajectories between Northern talkers and New England and Mid-Atlantic talkers. As shown in Figure 16, the trajectory of [ʊ] for the Northern dialect is fronted and slightly lower, whereas the trajectory for the other dialects is fronted with either little change in F1 or slight raising. The trajectory of the Northern talkers essentially crosses that of the Mid-Atlantic and New England talkers over the course of the vowel.

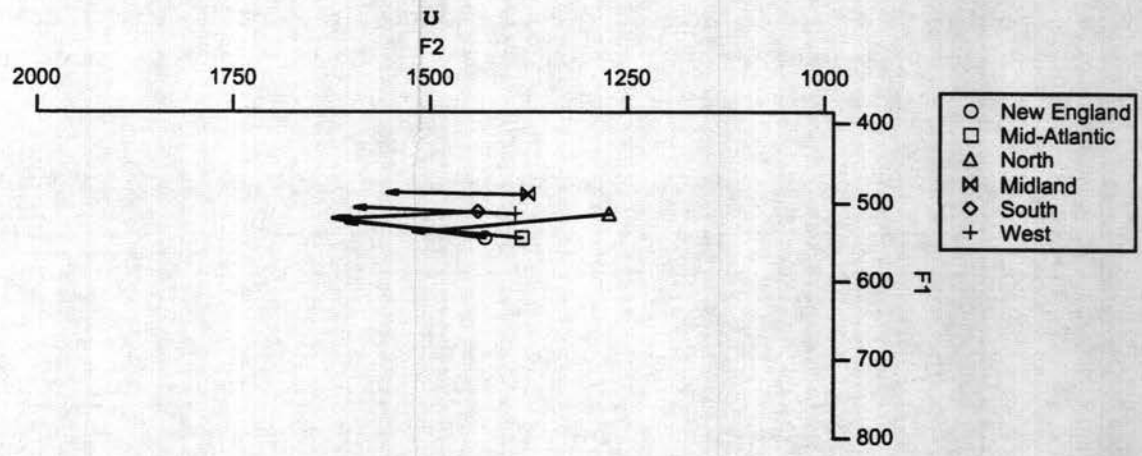


Figure 16. Mean formant frequency trajectories for each dialect for the vowel [ʊ].

For ΔF2, significant main effects of dialect were found for [ɪ] ($F(5, 234) = 4.4$, $p = .001$), [ɛ] ($F(5, 228) = 7.2$, $p < .001$), [æ] ($F(5, 233) = 12.8$, $p < .001$), [ʌ] ($F(5, 229) = 10.1$, $p < .001$), [o] ($F(5, 232) = 6.8$, $p < .001$), and [ɑw] ($F(5, 186) = 7.8$, $p < .001$).[7] As with the ΔF1 results discussed above, some of these differences are due to compensation for vowel shifts. Post-hoc Tukey tests revealed significant differences in ΔF2 in [ʌ] between Northern talkers and New England, Southern, and Western talkers (all $p < .001$) and in [æ] between Northern talkers and Mid-Atlantic, Midland, and Western talkers (all $p < .001$). Figure 17 shows the mean formant frequency trajectories for [ʌ] for each of the dialect groups. The greater ΔF2 exhibited by the Northern talkers is compensation for an initially backed vowel as part of the Northern Cities Chain Shift. This finding is similar to the Northern talkers' greater backward movement in F2 of [æ], which compensates for the Northern Cities Chain Shift raising and fronting of [æ], as discussed above and shown in Figure 14.

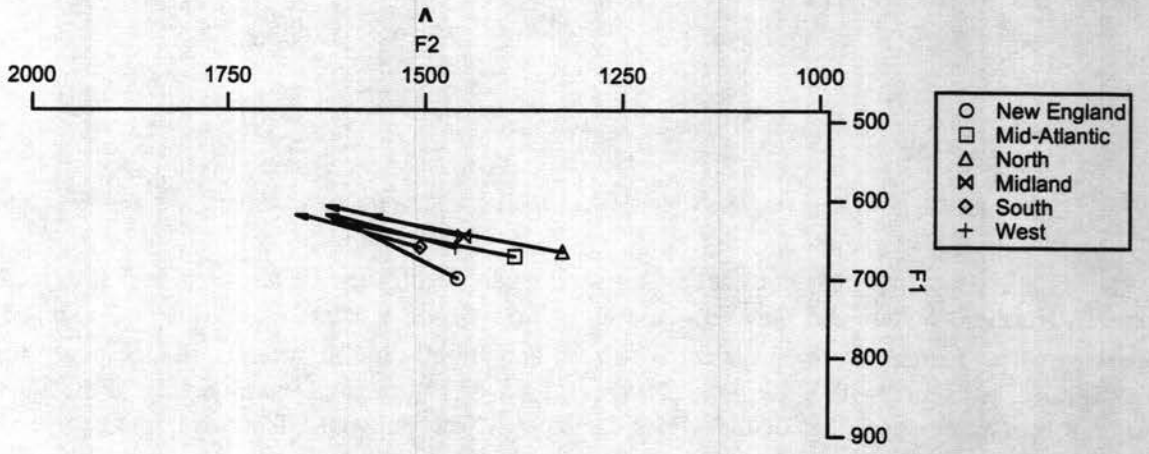


Figure 17. Mean formant frequency trajectories for each dialect for the vowel [ʌ].

[7] Increasing the p-value to .01 leads to two additional significant results for the ΔF2 analysis. Dialect differences were significant for [ey] and [ʊ] at $p < .01$.

Post-hoc Tukey tests also revealed significant differences in ΔF2 between Northern talkers and Midland and Southern talkers for [ɛ]. As discussed above and shown in Figure 15, this difference is due primarily to an overall difference in formant trajectory direction, with the Northern [ɛ] moving forward and the Midland and Southern [ɛ]s moving back. These differences in [ɛ] trajectories for Northern and Southern talkers are compensation for the Northern Cities Chain Shift [ɛ] backing and the Southern Vowel Shift [ɛ] fronting and raising.

A general difference in formant trajectory direction is responsible for the significant differences revealed by post-hoc Tukey tests between the Mid-Atlantic talkers and the New England and Northern talkers for ΔF2 of [o] (both $p < .001$). As can be seen in Figure 18, the Mid-Atlantic [o] moves forward, whereas the other dialects all show backing over the course of the vowel.

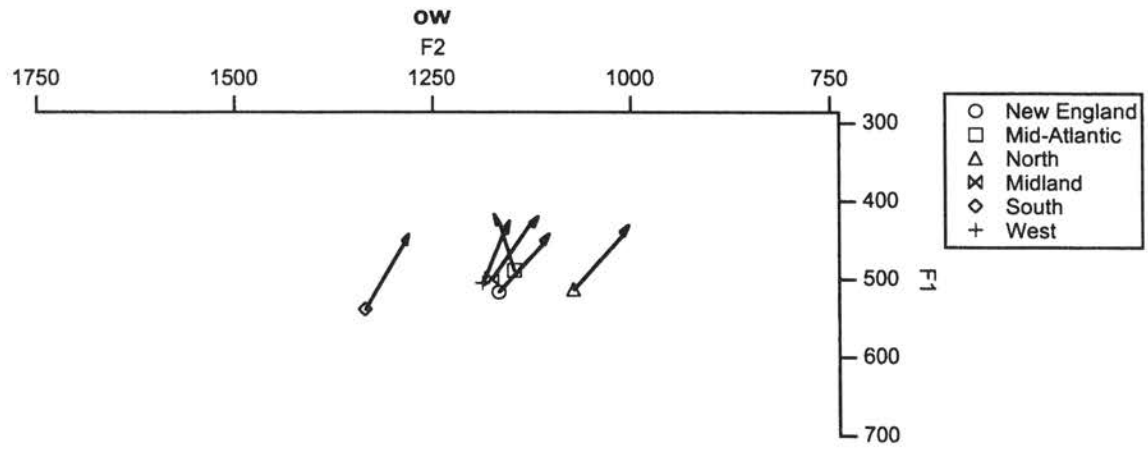


Figure 18. Mean formant frequency trajectories for each dialect for the vowel [o].

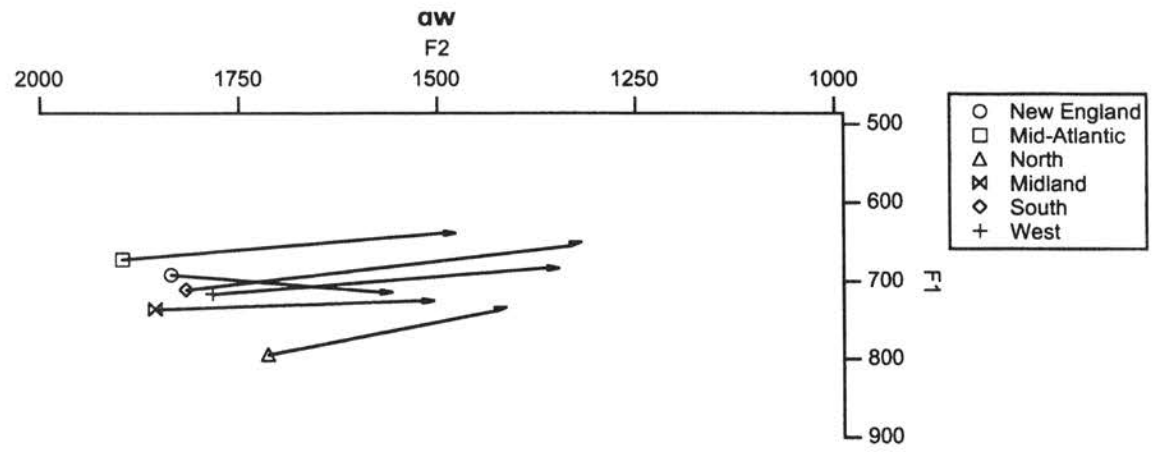


Figure 19. Mean formant frequency trajectories for each dialect for the vowel [aw].

Post-hoc Tukey tests on ΔF2 for [aw] revealed a significant difference between Southern talkers and New England and Northern talkers (both $p < .001$). The interpretation of these results is clear from the plot of the formant trajectories of [aw] in Figure 19. Southern talkers simply show

greater diphthongization of this vowel than New England and Northern talkers. Post-hoc Tukey tests also revealed a significant difference between the Western and Southern talkers for ΔF2 in the vowel [ɪ]. The mean formant trajectories for each dialect for [ɪ] are shown in Figure 20. Once again, the difference in formant trajectory in F2 is due to greater movement in the Southern productions of [ɪ] than the Western variants. Thomas (2001) reported that [ɪ] ingliding often accompanies [ɪ] fronting in the Southern Vowel Shift, although we do not have strong evidence of [ɪ] fronting in these data.

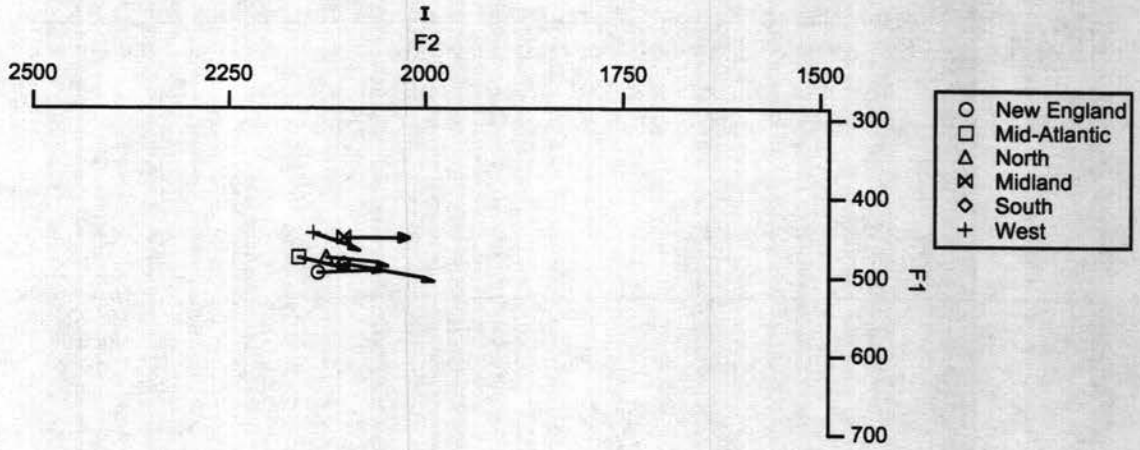


Figure 20. Mean formant frequency trajectories for each dialect for the vowel [ɪ].

Merger of [ɑ] and [ɔ]. To assess the degree of merger of [ɑ] and [ɔ], a series of paired-sample t-tests was calculated. For each of the six dialect regions, one paired-sample t-test was computed for F1 and one was computed for F2. Significant differences in this analysis suggest distinct vowels, whereas non-significant differences in both F1 and F2 suggest a merger. Due to the large number of comparisons, the p-value was set at .005 for this analysis. A merger of [ɑ] and [ɔ] was found for the Mid-Atlantic talkers ($t(7) = 2.8$, $p = .03$ for F1 and $t(7) = 2.2$, $p = .06$ for F2), the Midland talkers ($t(7) = 2.4$, $p = .04$ for F1 and $t(7) = 2.1$, $p = .07$ for F2), and the Western talkers ($t(7) = 3.5$, $p = .01$ for F1 and $t(7) = -.22$, $p = .83$ for F2). The New England talkers exhibited a near-merger of [ɑ] and [ɔ] ($t(7) = 5.5$, $p = .001$ for F1 and $t(7) = 3.7$, $p = .007$ for F2). Finally, [ɑ] and [ɔ] were clearly distinct for the Northern talkers ($t(7) = 7.6$, $p < .001$ for F1 and $t(7) = 7.1$, $p < .001$ for F2) and the Southern talkers ($t(7) = 6.7$, $p < .001$ for F1 and $t(7) = 4.9$, $p = .002$ for F2).[8]

Discussion

The statistical analysis of the acoustic vowel duration and formant frequency measures confirmed the presence of the Northern Cities Chain Shift in the Northern talkers and the Southern Vowel Shift in the Southern talkers. In particular, Northerners produced lowered and fronted [ɑ]s, fronted [æ]s, and backed [ʌ]s. In addition, the analysis of the formant trajectories suggested backed productions of [ɛ] for the Northern talkers. The Southern talkers exhibited fronting of [u] and [ow] as well as [ey] lowering. [ɪ] fronting by the Southern talkers was also suggested by greater F2 movement over the course of the vowel for the Southerners than for the other talkers. The Southerners also

[8] Increasing the p-value to .01 suggests a distinction for the New England talkers and a near-merger for the Western talkers.

produced longer lax vowels than any of the other talkers. Thus, the most robust dialect differences in the corpus are the Northern Cities Chain Shift among the Northern talkers and the Southern Vowel Shift among the Southern talkers.

The data plotted for each dialect in Figures 6 through 11 and for each talker in Appendix B suggest other important dialect-specific features are present in at least some talkers, although they might not be strong enough to be significant in the statistical analyses. For example, some of the Westerners showed the predicted [u] fronting while some did not. In particular, the Western talkers from Montana, Nevada, and one from California did not exhibit [u] fronting, while the other five Western talkers did. The Westerners also produced a merger of [ɑ] and [ɔ], as predicted.

The low-back merger was also reliable in the Midland talkers. In addition, some Midland talkers showed Southern features, such as [u] fronting while others showed Northern features such as [æ] raising. All three of the talkers who exhibited these Northern and/or Southern features in their speech are from the Indianapolis metropolitan area. These results are consistent with earlier claims in the literature that the Midland dialect region is not a unique dialect, but instead may be a transition area between the North and the South (Davis & Houck, 1992, but see also Frazer, 1994; Johnson, 1994).

Boberg (2001) claimed that Western New England can also be treated as a transition area between Eastern New England and the North. At first glance, the data presented here support this interpretation: some of the New England talkers showed the [æ] raising found in Western New England and some did not. However, [æ] raising among the male talkers was found in both Eastern and Western New England, whereas females from Western New England were split on [æ] raising with one producing the raised variant (NE8) and one not (NE7). In addition, all of the New England talkers, both Easterners and Westerners, showed the merger of [ɑ] and [ɔ] reportedly found in Eastern New England. A near-merger of [ɑ] and [ɔ] in the New England talkers was confirmed by the statistical analysis. These results suggest that Eastern and Western New England are perhaps more homogeneous than previously suggested (Boberg, 2001; Labov et al., forthcoming).

The distinctive features of the Mid-Atlantic talkers were less clear. We did not find evidence of a raised [ɔ] in most of the speakers, and instead, found evidence of an unexpected merger of [ɑ] and [ɔ]. Finally, [u] fronting also occurred in some Mid-Atlantic talkers but not others. In general, although these talkers are impressionistically unmistakable as Mid-Atlantic natives, neither the statistical analysis nor the examination of the vowel plots revealed interpretable acoustic-phonetic properties to define these talkers. Further research is needed to explore the role of other sources of dialect-specific variation, such as consonantal and prosodic differences, that may distinguish the Mid-Atlantic dialect from other regional varieties of American English.

Two important assumptions underlie this discussion of the different vowel systems of regional varieties of American English. First, the characteristics of each dialect are defined relative to an unspecified baseline. This baseline could be defined historically in terms of earlier vowel systems in the United States. For example, Labov (1994) characterized the Northern Cities Chain Shift in historical terms, by describing the vowel system of the Northern dialect as the result of a series of phonological changes that can be traced through both real-time and apparent-time data over the course of the second half of the 20th century. An alternative to this historical perspective would be to identify baseline pronunciations based on the current vowel systems of the different regional varieties. For example, Figure 5 showed highly consistent productions of [i] across the six different dialects, but

much greater variation in [æ], [ɑ], and [ow]. In the case of [æ], the Northern talkers were the only ones who produced the raised and fronted variant, suggesting that the lowered and backed production should be treated as the baseline. Similarly, the Southern [ow] was fronted, while a backed [ow] was found in the other five dialects, suggesting that the backed variant should be treated as the baseline. By considering both historical developments and synchronic idiosyncrasies, sociolinguists have developed a set of implicit baseline productions to which many other possible variants are compared (Thomas, 2001).

Second, despite significant acoustic-phonetic differences in the production of a single vowel category across the different dialects, we assume that the phonemic inventory of the different regional varieties included in the NSP corpus are essentially identical (Thomas, 2001). That is, Figures 5 through 11 depict 11 monophthongal vowel categories and the statistical analyses evaluated the productions of 14 vowel categories. With the exception of the merger of [ɑ] and [ɔ], which reduces the vowel inventory by one phoneme, none of the other vowel shifts or variants described above are assumed to affect the inventory of vowel phonemes that naïve listeners perceive or produce.

Finally, while talker gender was not explicitly examined in the present study, it should be noted that in some cases it appears that the male talkers were actually more advanced in certain sound changes than the female talkers (e.g., the Midland male talkers showed some Northern and some Southern features, whereas the Midland female talkers did not). This trend was unexpected given that women are typically thought to lead men in phonological change (Labov, 1990, 2001). A complete analysis of the interaction between gender and dialect in vowel production is necessary to confirm this apparent trend. An analysis involving gender may also provide additional insights into the characteristic properties of the speech of the talkers from different regions in the NSP corpus.

Conclusions

The Nationwide Speech Project is a new corpus containing recordings of 60 young adult talkers representing six different regional varieties of American English: New England, Mid-Atlantic, North, Midland, South, and West. The speech samples obtained from each talker include isolated words, isolated sentences, passages, and interview speech. The results of the acoustic analysis reported above confirm the presence of significant differences in vowel production due to dialect in the NSP corpus. The differences uncovered by the statistical analysis are robust across gender and should therefore be available to naïve listeners in perceptual dialect classification tasks, such as those described in Chapters 3 and 4.

References

Ash, S. (n.d.) Sampling strategy for the Telsur/Atlas project. http://www.ling.upenn.edu/phono_atlas/sampling.html.

Baker, K., Ramig, L., Johnson, A., & Freed, C. (1997). Preliminary voice and speech analysis following fetal dopamine transplants in five individuals with Parkinson disease. *Journal of Speech, Language, and Hearing Research, 40*, 615-626.

Boberg, C. (2001). The phonological status of Western New England. *American Speech, 76*, 3-29.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication, 15*, 39-54.

Canavan, A., & Zepperlen, G. (1996a). CALLFRIEND American English-Non-Southern Dialect (LDC96S46). Philadelphia: Linguistic Data Consortium.

Canavan, A., & Zepperlen, G. (1996b). CALLFRIEND American English-Southern Dialect (LDC96S47). Philadelphia: Linguistic Data Consortium.

Carter, A.K., & Clopper, C.G. (2002). Prosodic effects on word reduction. *Language and Speech, 45*, 321-353.

Carver, C.M. (1987). *American Regional Dialects: A Word Geography*. Ann Arbor, MI: University of Michigan Press.

Clopper, C.G., Carter, A.K., Dillon, C.M., Hernandez, L.R., Pisoni, D.B., Clarke, C.M., Harnsberger, J.D., & Herman, R. (2001). The Indiana Speech Project: An overview of the development of a multi-talker multi-dialect speech corpus. In *Research on Spoken Language Processing Progress Report No. 25* (pp. 367-380). Bloomington, IN: Speech Research Laboratory, Indiana University.

Clopper, C.G., Conrey, B.L., & Pisoni, D.B. (in press). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology*.

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.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1975). *Motor Speech Disorders*. Philadelphia: Saunders.

Davis, L.M., & Houck, C.L. (1992). Is there a Midland dialect? – Again. *American Speech, 67*, 61-70.

Dictionary of American Regional English. (1985-). Volume 1 (A-C), F.G. Cassidy (Ed.), Volume 2 (D-H) and Volume 3 (I-O), F.G. Cassidy & J.H. Hall (Eds.), Volume 4 (P-Sk), J.H. Hall (Ed.). Cambridge, MA: Belknap Press.

DuBois, J.W., Chafe, W.L., Meyer, C., & Thompson, S.A. (2000). Santa Barbara Corpus of Spoken American English Part-I (LDC2000S85). Philadelphia: Linguistic Data Consortium.

Fairbanks, G. (1940). *Voice and Articulation Drillbook*. New York: Harper.

Feagin, C. (2002). Entering the community: Fieldwork. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 20-39). Malden, MA: Blackwell.

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

Frazer, T.C. (1994). On transition areas and the 'Midland' dialect: A reply to Davis and Houck. *American Speech, 69*, 430-435.

Gelfer, M., & Schofield, K. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice, 14*, 22-33.

Giles, H., & Powesland, P. (1997). Accommodation theory. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A Reader and Coursebook* (pp. 232-239). New York: Palgrave.

Hall, J.H., & von Schneidemesser, L. (2004). Dictionary of American Regional English. http://polyglot.lss.wisc.edu/dare/dare.html.

Hillenbrand, J., & Houde, R. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research, 39*, 311-321.

Johnson, E. (1994). Yet again: The Midland dialect. *American Speech, 69*, 419-430.

Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Krapp, G.P. (1925). *The English Language in America*. New York: Frederick Ungar.

Labov, W. (1972a). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1972b). Some principles of linguistic methodology. *Language in Society, 1*, 97-120.

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. (1994). *Principles of Linguistic Change: Internal Factors*. Malden, MA: Blackwell.

Labov, W. (1998). The three dialects of English. In M.D. Linn (Ed.), *Handbook of Dialects and Language Variation* (pp. 39-81). San Diego: Academic Press.

Labov, W. (2001). *Principles of Linguistic Change: Social Factors*. Malden, MA: Blackwell.

Labov, W., Ash, S., & Boberg, C. (forthcoming). *Atlas of North American English*. New York: Mouton de Gruyter.

McDavid, R.I., Jr. (1958). The dialects of American English. In W.N. Francis (Ed.), *The Structure of American English* (pp. 480-543). New York: Ronald Press.

McHenry, M. (1999). Aerodynamic, acoustic and perceptual measures of nasality following traumatic brain injury. *Brain Injury, 13*, 281-290.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1994). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10* (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.

Plichta, B., & Mendoza-Denton, N. (2001). Best practices in the digital acquisition and processing of acoustic speech data. Paper presented at New Ways of Analyzing Variation 30, Raleigh, NC. October 11-13.

Rischel, J. (1992). Formal linguistics and real speech. *Speech Communication, 11*, 379-392.

Rojas, D.M. (2002). *PIQUANT-LA: Probabilistic Identification and Quantification of Linguistic Affinity*. Master's thesis, University of Edinburgh, Edinburgh, Scotland.

Sapienza, C., Walton, S., & Murry, T. (1999). Acoustic variations in adductor spasmodic dysphonia as a function of speech task. *Journal of Speech, Language, and Hearing Research, 42*, 127-140.

Sjölander, K., & Beskow, J. (2004). WaveSurfer (Version 1.6.2) [Computer software]. Stockholm: Centre for Speech Technology, Royal Institute of Technology.

Stockwell, P. (2002). *Sociolinguistics: A Resource Book for Students*. London: Routledge.

Strassel, S., Conn, J., Wagner, S.E., Cieri, C., Labov, W., & Maeda, K. (2003). SLX Corpus of Classic Sociolinguistic Interviews (LDC2003T15). Philadelphia: Linguistic Data Consortium.

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

Trager, G.L. (1930). The pronunciation of 'short a' in American Standard English. *American Speech, 5*, 396-400.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America, 88*, 97-100.

Trudgill, P. (1998). Accommodation between dialects. In M.D. Linn (Ed.), *Handbook of Dialect and Language Variation* (pp. 307-342). San Diego: Academic Press.

Wells, J.C. (1982). *Accents of English 3: Beyond the British Isles*. Cambridge: Cambridge University Press.

Wolfram, W., & Fasold, R.W. (1997). Field methods in the study of social dialects. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A Reader and Coursebook* (pp. 89-115). New York: Palgrave.

Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication, 9*, 351-356.

CHAPTER 3: FORCED-CHOICE PERCEPTUAL DIALECT CATEGORIZATION

Introduction

Dialect Categorization

The perception of dialect variation by naïve listeners is a growing research area in the field of sociolinguistics. Dialect geographers and variationist sociolinguists have been documenting regional and social linguistic varieties of American English for more than a century (McDavid, 1958) and the implications of variation for social interactions have been explored by social psychologists for several decades (Ryan & Giles, 1982). More recently, traditional speech science methods have been applied to the study of linguistic variation to uncover the categories that naïve listeners have for regional and social dialects (Thomas, 2002).

In one of the earliest studies of dialect categorization, Preston (1993) asked naïve adults to listen to a set of nine male talkers and then select the city that they thought each talker was from. The talkers were from nine different cities in the United States between Saginaw, Michigan and Dothan, Alabama. The listeners heard a short extract from an interview with each of the talkers and were asked to select from the nine cities the one that they thought the talker was from. Preston (1993) found that the naïve listeners could make reliable distinctions between Northern and Southern talkers, but that their ability to distinguish talkers from neighboring cities was more limited.

More recently, Clopper and her colleagues (Clopper, Conrey, & Pisoni, in press; Clopper & Pisoni, 2004b) also found that naïve listeners are able to make broad dialect categorizations. Clopper and Pisoni (2004b) played sentence-length utterances read by male talkers from six different dialect regions in the United States to naïve listeners. The stimulus materials were taken from the TIMIT Acoustic-Phonetic Continuous Speech Corpus which contains recordings of 630 talkers (Fisher, Doddington, & Goudie-Marshall, 1986). The TIMIT talkers include both males and females with a range of ages, ethnicities, and regional backgrounds. The six dialect regions examined by Clopper and Pisoni (2004b) were New England, North, North Midland, South Midland, South, and West. They found that naïve listeners were 31% accurate in categorizing unfamiliar talkers by dialect. While this performance was poor, it was statistically above chance in a six-alternative task. Clopper and Pisoni (2004b) also analyzed the pattern of errors produced in the categorization task and found that the listeners made systematic confusions between phonologically similar dialects. In particular, the listeners appeared to make consistent distinctions between Northeastern, Southern, and Western varieties of American English, but were far less accurate in identifying the six regional varieties used in the task. Clopper et al. (in press) replicated the earlier results reported by Clopper and Pisoni (2004b) for a group of female talkers and a mixed group of both male and female talkers, suggesting that categorization of unfamiliar talkers by dialect is reliable across both genders and in mixed gender conditions. Like Preston's (1993) earlier study, the experiments by Clopper and her colleagues revealed that naïve listeners are able to make some broad classifications about the region of origin of unfamiliar talkers.

In a study examining the effects of a single phonetic variable on dialect classification, Plichta and Preston (2003) asked naïve listeners to assign talkers to the same set of nine cities used in Preston's (1993) study using synthesized utterances of a single word. They created a continuum of seven stimulus items that exhibited a range of [ɑy] monophthongization using the word *guide*. They asked naïve participants to listen to each token and select the city that they thought the talker was from. The results revealed reliable categorizations of all seven of the stimulus tokens, with the more diphthongal tokens assigned to more Northern cities and the more monophthongal tokens assigned to

more Southern cities in a one-to-one correspondence from North to South. Significant differences were found between all neighboring tokens on the continuum in terms of their average geographic placement. These results suggest that naïve listeners have some knowledge about at least one of the linguistic variables that distinguish Northern from Southern talkers in the United States, [ɑy] monophthongization.

More recently, Clopper and Pisoni (2004b) measured some of the acoustic-phonetic properties of the speech signal that may have been salient to the naïve listeners in their categorization task. Following an acoustic analysis of the stimulus materials, Clopper and Pisoni (2004b) conducted a series of regression analyses to determine which phonetic variables were associated with the observed dialect affiliation of the talkers and which were related to the naïve listeners' judgments of dialect category. The results of the regression analysis revealed four main variables that served as reliable cues to dialect categorization for the naïve listeners: New England r-lessness in the word *dark*, New England backing of [æ] in the word *rag*, Northern [ow] offglide centralization in the word *don't*, and South Midland [u] fronting in the word *suit*. These findings suggest that naïve listeners have reliable sociolinguistic categories for phenomena such as regional dialects and that these categories are related to phonological and phonetic variation between those dialects.

Dialect categorization studies have also been carried out with naïve listeners in the Netherlands and the United Kingdom. Williams, Garrett, and Coupland (1999) recorded narratives in English from two adolescent male talkers from each of six different regions of Wales, plus two adolescent male speakers of Received Pronunciation (RP). They played 30-second extracts of these narratives to different adolescent males in Wales and asked them to select the region that they thought each talker was from in an eight-alternative forced-choice categorization task (the response categories were the six regions of Wales, RP, and "don't know"). They found that the children performed the task with overall accuracy of approximately 30%.

In another study, Van Bezooijen and Gooskens (1999) found much higher performance for residents of the United Kingdom in a dialect categorization task. Using brief narratives from three male talkers from each of five regions in the United Kingdom, Van Bezooijen and Gooskens (1999) found that adult listeners could identify the country, region, and area of origin of unfamiliar talkers with 92%, 88%, and 52% accuracy, respectively. The discrepancy in performance between the Welsh adolescents and the British adults may be due to the relatively greater geographic and linguistic differences between the talkers in the Van Bezooijen and Gooskens (1999) task, who were from all over the United Kingdom, than the talkers in the Williams et al. (1999) study, who were all from Wales.

Van Bezooijen and Gooskens (1999) also conducted a similar study in the Netherlands, using three male talkers from each of four regions in the Netherlands and Dutch-speaking Belgium. They found that adult listeners in the Netherlands could identify the country, region, and province of origin of the talkers with 90%, 60%, and 40% accuracy, respectively. These results were replicated for female Dutch speakers using read speech materials instead of narratives by Van Bezooijen and Ytsma (1999).

Taken together, these dialect categorization studies suggest that naïve listeners are able to identify unfamiliar talkers by regional dialect with above-chance performance. Although performance varies across cultures, with the lowest performers being Americans, none of the groups of listeners were performing with exceptionally high levels of accuracy. Despite these relatively poor levels of performance, however, the results of the regression analysis by Clopper and Pisoni (2004b) and the identification task using synthetic speech tokens by Plichta and Preston (2003) suggest that naïve

listeners were attending to some of the relevant properties in perception that distinguished the different regional varieties. That is, the listeners were relying in part on reliable acoustic-phonetic differences between dialects when performing these kinds of explicit dialect categorization tasks.

Additional research on the perceptual categorization of dialect variation is clearly needed, particularly in the United States. Preston's (1993) identification study was limited to a North-South continuum and did not explore any variation in the East-West dimension. Similarly, the use of the TIMIT corpus by Clopper and her colleagues (Clopper et al., in press; Clopper & Pisoni, 2004b) raises some concerns about the interpretation of the results. The TIMIT corpus was originally designed for use in speech recognition research (Fisher et al., 1986) and as a result, the sociolinguistic properties of the corpus were not strictly controlled. For example, the original regional labels used to describe the dialects of the talkers were questionable, both in terms of their names and the geographic regions that they covered. In addition, it is unclear what criteria were used to determine which region a given talker should be assigned to and specific information about the residential history of the talkers is no longer available. The current study therefore used methods similar to those in the earlier work by Clopper and her colleagues (Clopper et al., in press; Clopper & Pisoni, 2004b), but the stimulus materials were taken from the new Nationwide Speech Project corpus which was designed specifically for use in perceptual and acoustic studies of regional variation in the United States (Chapter 2).

Residential History and Dialect Categorization

Some recent findings on perceptual dialect categorization suggest that the linguistic experience and residential history of the listeners may affect performance. For example, Preston (1993) conducted his dialect identification task in Michigan and southern Indiana. He found that the overall performance of the two groups was similar, because both groups appeared to distinguish between Northern and Southern talkers. However, the geographic boundary between the North and South was different for the two groups. In particular, the Michigan listeners heard the major North-South boundary between Indiana and Kentucky, whereas the Indiana listeners heard the major North-South boundary between Kentucky and Tennessee. Preston (2002) attributed this difference in perception to differences in "linguistic security" in the two regions.[9] In particular, Michigan listeners are typically found to be linguistically secure, whereas listeners in southern Indiana are typically less linguistically secure (Preston, 1993, 2002). Preston (2002) argued that this difference in perceived prestige of their own variety led to different identification strategies. In particular, the southern Indiana listeners used cues to perceived "pleasantness" in making their categorization judgments, whereas the listeners from Michigan relied on cues to perceived "correctness."

In their study of dialect categorization in Wales, Williams et al. (1999) also found several effects of residential history and linguistic experience on performance. First, although the performance of the adolescent boys was 30% correct overall, a more detailed examination of the results revealed that the listeners were far more accurate in the categorization of talkers from their own region (45%) than of talkers from other regions (24%). Second, Williams et al. (1999) also asked schoolteachers in Wales to participate in the dialect categorization task and found that the adults were more accurate than the children, with overall accuracy of 52% across all of the talkers for the adults, compared to 30% overall accuracy for the children. Williams et al. (1999) attributed this difference in categorization performance between the children and adults to the greater travel experiences of the adults as compared to the children. Thus, these two additional findings from the Welsh study suggest

[9] Linguistic security is defined with respect to the participants' ratings of the "correctness" of their speech and the speech of others. Linguistically secure participants rate their own speech as highly correct. Linguistically insecure participants rate their own speech as less correct than some other varieties which are perceived to be the standard or norm.

that exposure to different regional varieties through residency in certain regions as well as travel experience may have effects on dialect categorization accuracy.

Clopper and Pisoni (2004a) explicitly examined the effects of residential history on dialect categorization performance using the same six-alternative forced-choice task described above. Two groups of participants were recruited who differed in their residential history with respect to geographic mobility. One group, the "mobile" listeners had lived in at least three different states in the United States. The second group, the "non-mobile" listeners, had lived only in Indiana. We found that the mobile group performed slightly better than the non-mobile group, but this difference was only significant for one of the three blocks of trials. A more detailed analysis of the performance by the mobile listeners with respect to which dialect regions they had lived in revealed more interesting results. In particular, listeners who had lived in a given region ("residents") performed more accurately on talkers from that same region than listeners who had not lived there ("non-residents"). This result was robust across all of the dialect regions.

Two recent studies have also shown that linguistic experience and residential history can affect listeners' ability to perceptually adapt to the dialect of an unfamiliar talker. In the United States, Rakerd and Plichta (2003) examined the perception of the Northern Cities Chain Shift by listeners from Detroit and the Michigan Upper Peninsula. The listeners were asked to identify a series of synthetic stimulus items as *sock* or *sack*. The stimuli varied in their second formant frequency and were presented to the listeners both in isolation and at the end of carrier sentences produced by a Detroit talker and an Upper Peninsula talker. The Detroit talker produced the shifted vowels found in the Northern dialect region, whereas the Upper Peninsula talker did not. Rakerd and Plichta (2003) found that the Detroit listeners, who were exposed to the Northern Cities Chain Shift in their local variety, were able to adapt to the dialect of the talker producing the carrier sentence and selected *sock* for more stimuli with higher second formant frequencies for the Detroit carrier sentence than for the Upper Peninsula carrier sentence or the words in isolation. However, the listeners from the Upper Peninsula, who were not as familiar with the characteristic vowel productions of the Northern Cities Chain Shift, appeared to use the same second formant frequency cutoff for distinguishing *sock* from *sack*, regardless of the context.

In a similar study in Great Britain, Evans and Iverson (2004) asked listeners in Northern and Southern Britain to rate synthetic stimuli embedded in carrier sentences on their goodness as exemplars of Northern or Southern British English. They found that listeners from Northern England with little exposure to Southern varieties showed less dialect-specific variation in their responses and tended to select Northern variants regardless of the target dialect. In contrast, listeners from Northern England who had moved to Southern England showed greater adaptation to the dialect in the carrier sentence and selected more accurate variants for the Southern British vowels. The studies by Evans and Iverson (2004) and Rakerd and Plichta (2003) suggest that prior linguistic experience can affect the perception of the phonological properties that distinguish different dialects.

Finally, one early study also revealed a relationship between residential history and cross-dialect intelligibility. Mason (1946) reported the results of a speech intelligibility experiment in noise in which the talkers and listeners came from different regions of the United States. He found that performance was better when the talkers and listeners shared a dialect than when the talkers and listeners were from different dialect regions. These findings suggest that residential history affects the perception of the linguistic content of the message, in addition to the dialect of talker.

Taken together, the results of these studies suggest that experience with language variation affects performance on explicit tasks of dialect categorization and identification, as well as speech intelligibility. In particular, Preston (1993), Rakerd and Plichta (2003), Williams et al. (1999), and

Mason (1946) found that geographic location can affect perception. In addition, Clopper and Pisoni (2004a), Evans and Iverson (2004), and Williams et al. (1999) found that exposure to different varieties through travel or geographic mobility can also lead to differences in perception. In general, geographic mobility leads to greater overall accuracy in explicit categorization and identification tasks, whereas the location of the listeners interacts with the dialect of the talkers and leads to more accurate categorization responses when the talkers and listeners have the same regional dialect than when they do not.

Early Linguistic Experience, Dialect Acquisition, and Dialect Maintenance

Early linguistic experience has been shown repeatedly to affect phoneme production and perception and a number of researchers have therefore suggested that some aspects of language acquisition may be susceptible to sensitive or critical periods in development (Bruer, 2001; Scovel, 1988). For example, while young infants can discriminate phonemes that do not contrast in their native or ambient language, children quickly become attuned to the relevant contrasts in their native language (Aslin & Pisoni, 1980; Jusczyk, 1997). Adults therefore often find it difficult to discriminate two phonemes in a foreign language that are non-contrastive in their native language (Polka, 1992, 1995). In addition, early exposure to non-native contrasts can have lasting perceptual effects. Tees and Werker (1984) reported that native English speakers who had been exposed to Hindi at a young age were able to accurately discriminate Hindi voiceless retroflexed and dental stops, whereas native English speakers who had not been exposed to Hindi before the age of 2 years were unable to make the discrimination until they had as many as five years of experience learning Hindi as adults. Thus, some aspects of phonological competence in the native language are acquired early in infancy and at the same time, phonological input during this early stage of language acquisition can have dramatic effects on speech perception and discrimination abilities in adulthood (Strange, 1995).

The critical period hypothesis is also discussed in relation to the acquisition of a second language. In particular, it is well-documented that age of acquisition of a second language can have a large impact on the degree to which the learner is able to acquire native-like pronunciation in the target language (Strange, 1995). For example, Flege, Munro, and MacKay (1995) found significant correlations between age of learning and overall accentedness ratings for second language learners of English. With respect to the perception of non-native contrasts, Yamada (1995) reported that Japanese learners of English exhibited age-related effects in the discrimination of English /r/ and /l/. In particular, the Japanese participants who had been in the United States for longer periods of time and started learning English at a younger age, performed more like native English speakers than those who had less experience with English. The results of these studies led many researchers to the conclusion that unaccented second language acquisition is virtually unattainable after puberty (e.g., Scovel, 1988, but see Hakuta, 2001).

Despite the extensive literature on first and second language acquisition, relatively little is known about the development and acquisition of sociolinguistic competence, with respect to the production and perception of stylistic and social variation (Roberts, 2002). It is typically reported that children learn the local linguistic variety, regardless of the native dialect or language of their parents (Chambers, 2002; Labov, 1971). In addition, once children have reached adolescence, peer groups play a significant role in determining sociolinguistic performance as children begin to assert their independence from their parents (Eckert, 1988).

Several studies in Philadelphia have revealed complex interactions between the geographic mobility of children and the mobility of their parents in the acquisition of local dialect features. For example, Payne (1980) reported that children who moved to the Philadelphia area did not completely

acquire the complex back-vowel system of the local variety. Moreover, the degree to which the children acquired the local variety was related to the age at which they moved to Philadelphia. The children who were the youngest when they arrived acquired more of the variables than the children who were older. Payne (1980) and Roberts (1997) also found that children who were born in Philadelphia but whose parents were raised somewhere else were somewhat less successful in acquiring the local dialect than children whose parents were from Philadelphia. The results of these studies suggest that both the dialect of the parents and the local dialect affect first dialect acquisition.

The relationship between a critical period for language development and second dialect acquisition remains unclear, however. For example, Munro, Derwing, and Flege (1999) examined the acquisition of Southern American English in Alabama by native speakers of Canadian English and found that the speech of Canadians in Alabama was perceived as being distinct from both native Canadian English and native Alabama English. However, Chambers (1992) found that native Canadian children who moved to southern England acquired very few of the local phonological features. Finally, Bowie (2000) reported that adults who had moved out of the Waldorf, Maryland area revealed significant differences in both production and perception of certain vowels compared to adults who were lifetime residents of Waldorf. However, Bowie (2000) suggested that geographic mobility may lead to dialect attrition instead of dialect acquisition. That is, speakers of a given dialect may have difficulty maintaining the characteristic properties of that dialect when they are no longer surrounded by other speakers of that dialect. For example, the lifetime residents of Waldorf exhibited a perceptual merger of [ʊ] and [u] before [l], so that *pull* and *pool* were perceived as homophones. The participants who had moved away from Waldorf typically perceived two distinct forms, however, despite the fact that some of them moved to other locations in the United States where the merger is common in perception and/or production (Labov et al., forthcoming). Therefore, instead of acquiring the distinctive variants of the local dialect, the listeners lost some of the characteristic properties of their first dialect (Markham, 1997).

Taken together, the research on dialect acquisition suggests that the development of sociolinguistic competence results from both linguistic experience and social motivation. In particular, while age of exposure seems to affect acquisition of some phonetic and phonological variables (Payne, 1980; Roberts, 1997), age does not seem to play a central role in the loss of native features as a result of mobility (Bowie, 2000; Munro et al., 1999). In addition, even in children, some of the research findings suggest that social factors such as the desire to maintain a native dialect or assert independence from their parents may be more important than a critical or sensitive period in determining degree of acquisition of a local dialect (Chambers, 1992; Eckert, 1988). Additional research in this area is needed to determine the nature of the parallels between dialect and language acquisition and the effects of age, linguistic experience, and social factors on the development of sociolinguistic competence (Politzer, 1993).

The present study was designed to explore in more detail the relationship between the linguistic experience of the listeners and dialect categorization performance. In particular, the geographic location and mobility of the listeners were manipulated to create four listener groups: Mobile Northerners, Mobile Midlanders, Non-Mobile Northerners, and Non-Mobile Midlanders. Dialect categorization performance was assessed using a six-alternative forced-choice categorization task that was similar to the experimental methods used by Clopper and her colleagues (Clopper et al., in press; Clopper & Pisoni, 2004a, b), except that the stimulus materials were taken from the new Nationwide Speech Project corpus (Chapter 2).

Methods

Listeners

One hundred and fifteen listeners aged 18-25 years old were recruited from the Indiana University community for participation in this study. Data from 16 participants were excluded prior to the data analysis for the following reasons: three performed the task consistently at chance,[10] nine knew one or more of the talkers by name, one was fluent in a language other than English, and three reported a history of a hearing or speech disorder at the time of testing. The remaining 99 listeners were monolingual native speakers of American English with native English-speaking parents and no reported hearing or speech disorders. The participants received $8 for their service.

The 99 listeners who participated in the current experiment were assigned to one of four groups, based on their residential history. The 25 listeners in the Non-Mobile Midland group had lived only in the Midland dialect region.[11] The 25 listeners in the Non-Mobile North group had lived only in the Northern dialect region, prior to attending college in Bloomington, Indiana and their parents still lived in the Northern dialect region at the time of testing. The 25 listeners in the Mobile Midland group had lived in at least one dialect region other than the Midland before the age of 18 years old and their parents lived in the Midland dialect region at the time of testing. The 24 listeners in the Mobile North group had lived in at least one dialect region other than the North before the age of 18 years old and their parents lived in the Northern dialect region at the time of testing. The four groups of listeners therefore represented two degrees of geographic mobility (mobile and non-mobile) and two geographic locations (North and Midland) as shown in Table 1.

Table 1. Residential history of the 99 listeners in the six-alternative forced-choice categorization experiment.

	North	**Midland**	**Total**
Mobile	24	25	49
Non-Mobile	25	25	50
Total	49	50	99

Talkers

Twenty-four male and 24 female talkers were selected from the Nationwide Speech Project (NSP) corpus (Chapter 2) for the current study. The talkers included four males and four females from each of the six dialect regions included in the corpus: New England, Mid-Atlantic, North, Midland, South, and West. The hometowns of each of the 48 talkers are shown in Figure 1. The dark circles indicate male talkers and the light squares indicate female talkers. Forty-eight of the 60 talkers included in the NSP corpus were selected for the present study in order to allow for two complete blocks of novel sentence trials using high probability sentences. By reducing the number of talkers to

[10] Participants who did not correctly categorize at least five tokens from one of the six dialects were identified as performing consistently at chance and their data were excluded. It is impossible to determine if these participants were unable to perform the task accurately or if they were simply not attending to the task, but this strict exclusion criteria ensures that the data analyzed reflect the best efforts of the participants.

[11] The Northern and Midland dialect regions were geographically identical to those used in the Nationwide Speech Project corpus (Chapter 2). Specifically, highway US-30 served as the primary divide between the Northern and Midland regions in Ohio, Indiana, and Illinois.

48, two stimulus items could be presented for each talker without repeating any of the sentences during the course of the experiment. The four talkers of each gender from each dialect were selected randomly. This set of talkers is the same as the set of talkers whose vowel productions were analyzed in Chapter 2.

The six dialect regions represented by the talkers in the NSP corpus differ with respect to their vowel systems (Chapter 2). The Northern dialect is characterized by the Northern Cities Chain Shift which involves the clockwise rotation of the low and low-mid vowels, beginning with the raising and fronting of [æ] (Labov, 1998). The Southern dialect is characterized by the Southern Vowel Shift which involves the fronting of the high and mid back vowels, the centralization of the front tense vowels, and the peripheralization of the front lax vowels (Labov, 1998). Southern speech also contains monophthongal [ɑy] and [oy] (Thomas, 2001). The Mid-Atlantic dialect includes raising of [ɔ] and a split in words containing [æ] such that some exhibit raising and some do not (Labov, 1994; Thomas, 2001). The New England, Midland, and Western dialects all exhibit the low-back merger of [ɔ] and [ɑ] (Labov, 1998). New England speech also includes some aspects of the Northern Cities Chain Shift (Boberg, 2001). Finally, Western speech is also characterized by the fronting of [u] (Labov, Ash, & Boberg, forthcoming; Thomas, 2001). Chapter 2 describes the vowel systems of the talkers included in the NSP corpus in more detail.

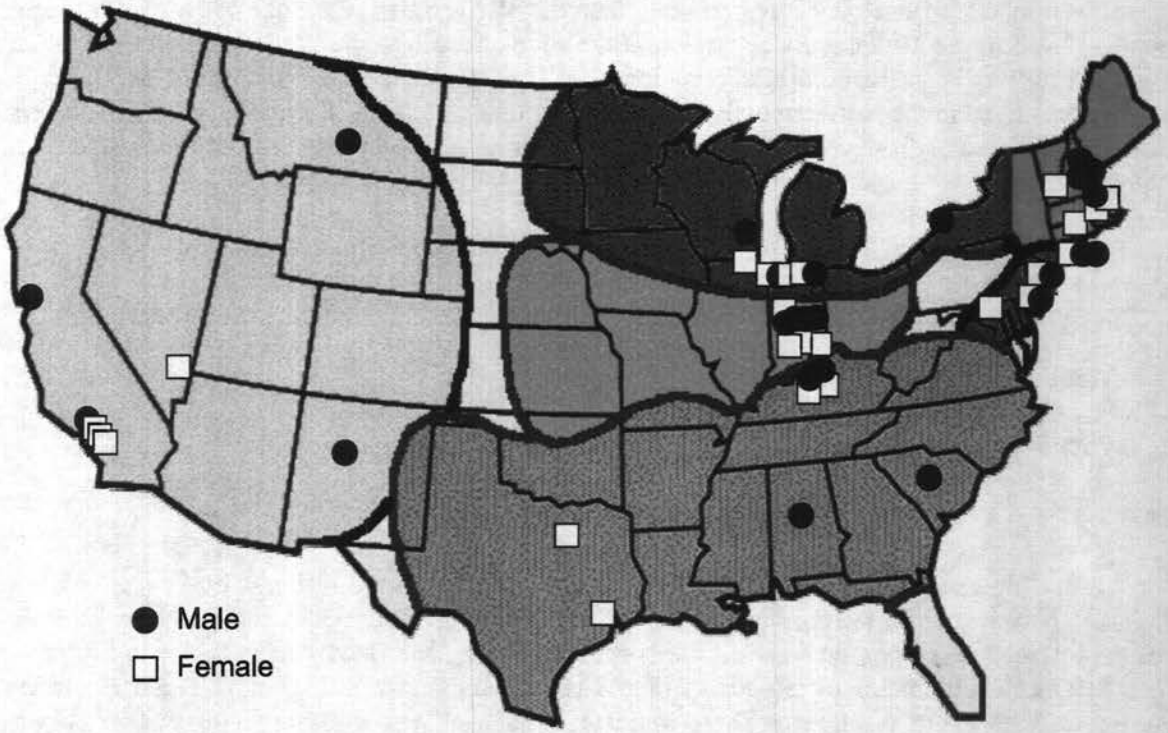


Figure 1. Map showing the hometowns of the 48 talkers in the six-alternative forced-choice categorization task. Dark circles indicate male talkers and light squares indicate female talkers.

Stimulus Materials

Two different high probability sentences were selected for each of the 48 talkers, for a total of 96 different stimulus items. The high probability sentences were taken from the Speech Perception in

Noise (SPIN) test and are meaningful English sentences that range in length from five to eight words. The final target word in each sentence is highly predictable from the preceding semantic content of the sentence (Kalikow, Stevens, & Elliot, 1977). The sentences were selected for each talker such that no sentence was repeated over the course of the experiment. Examples of the high probability sentences used in the present experiment are shown in (1). A complete list of the stimulus materials can be found in Appendix C.

(1) Ruth had a necklace of glass beads.
The swimmer dove into the pool.

The original digital sound files from the NSP corpus were edited to include only speech material and converted for presentation to the listeners to .wav digital sound files with 16-bit encoding and a sampling rate of 44.1kHz. In cases where multiple repetitions of a sentence were available in the NSP corpus, the final repetition was used because this token was most likely to be free of disfluencies or extraneous noise. The mean RMS amplitude level of each of the sound files was leveled to 67dB using Level16 (Tice & Carrell, 1998).

Procedure

Participants were seated at personal computers equipped with a mouse and Beyerdynamic DT100 headphones. The experiment consisted of two blocks of 48 trials each. In each block, participants heard one sentence from each of the 48 talkers one at a time in random order. For each talker, each of the two sentences was assigned randomly to either the first or second block to reduce stimulus-specific block effects. On each trial, the participants heard a single sentence over the headphones at approximately 70 dB SPL and were asked to select the region that they thought the talker was from. The response alternatives were displayed on a multi-colored map of the United States with a verbal label for each of the six regions, as shown in Figure 2. Before making their responses, the listeners were permitted to listen to each sentence as many times as they wanted by pressing a "Listen Again" button with the mouse. The listeners made their responses by pressing on the appropriate label on the screen. The experiment was self-timed and the listeners pressed a "Next Trial" button to proceed to the next trial. No feedback was provided to the listeners about the accuracy of their responses. Participants were permitted to take a break between the two blocks of trials.

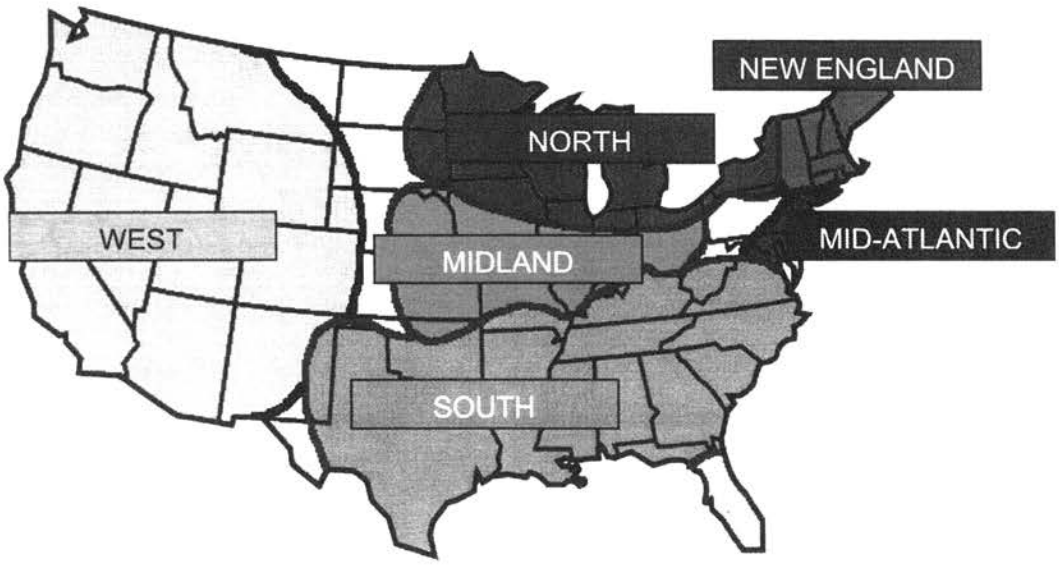


Figure 2. Response alternatives in the six-alternative forced-choice categorization task.

Results

Categorization Accuracy

A summary of the perceptual categorization performance for each of the four listener groups is shown in the middle column of Table 2. Chance performance in a six-alternative task is 17%. While the overall accuracy of the listeners was poor, all four groups were statistically above chance by a binomial test ($p < .05$).

Table 2. Overall mean percent correct performance and mean number of stimulus repetitions in the six-alternative forced-choice categorization task for each listener group, collapsed across experimental block and talker dialect. Standard deviations are shown in parentheses.

	Mean Percent Correct	Mean Stimulus Repetitions
Mobile North	27 (6)	1.65 (.49)
Mobile Midland	26 (7)	1.64 (.55)
Non-Mobile North	26 (8)	1.56 (.51)
Non-Mobile Midland	25 (5)	1.91 (1.24)
Total	26 (6)	1.69 (.77)

A repeated measures ANOVA with experimental block (first or second) and talker dialect (New England, Mid-Atlantic, North, Midland, South, or West) as within-subject variables and listener group (Mobile North, Mobile Midland, Non-Mobile North, or Non-Mobile Midland) as a between-subject variable revealed a significant main effect of talker dialect ($F(5, 475) = 43.2$, $p < .001$) and a significant block by dialect interaction ($F(5, 475) = 2.8$, $p = .015$). No other main effects or interactions were significant.

Figure 3 shows the percent correct performance across all four listener groups for each of the two experimental blocks for each of the six dialects. Post-hoc Tukey tests on talker dialect for each experimental block revealed the locus of the interaction as well as the overall main effect of dialect. For the first block of trials, performance on the New England talkers was worse than performance on any of the other talker groups (all $p < .05$). Performance on the Midland talkers was the best, with significant differences in performance revealed between the Midland talkers and all other groups except the South (all $p < .05$). Significant differences were also found between Mid-Atlantic and Western talkers ($p < .001$), Northern and Southern talkers ($p < .001$), and Southern and Western talkers ($p < .001$). Similarly, for the second block of trials, performance on the New England and Western talkers was significantly worse than performance on the other four talker groups (all $p < .05$). Performance on the Midland talkers was also best in the second block, with significant differences found between Midland talkers and all of the talker groups except the South (all $p < .01$). Overall, performance was best for Midland and Southern talkers and worst for New England and Western talkers, with performance on Mid-Atlantic and Northern talkers in between.

The locus of the block by dialect interaction reflects the improvement in performance on New England and Northern talkers in the second block relative to the first. Paired sample t-tests confirm a significant improvement in performance on New England and Northern talkers from the first to the second experimental block ($t(98) = -2.3$, $p < .05$ for New England and $t(98) = -3.0$, $p < .01$ for North). No significant differences in performance between the first and second experimental blocks were found for the other four talker dialect groups.

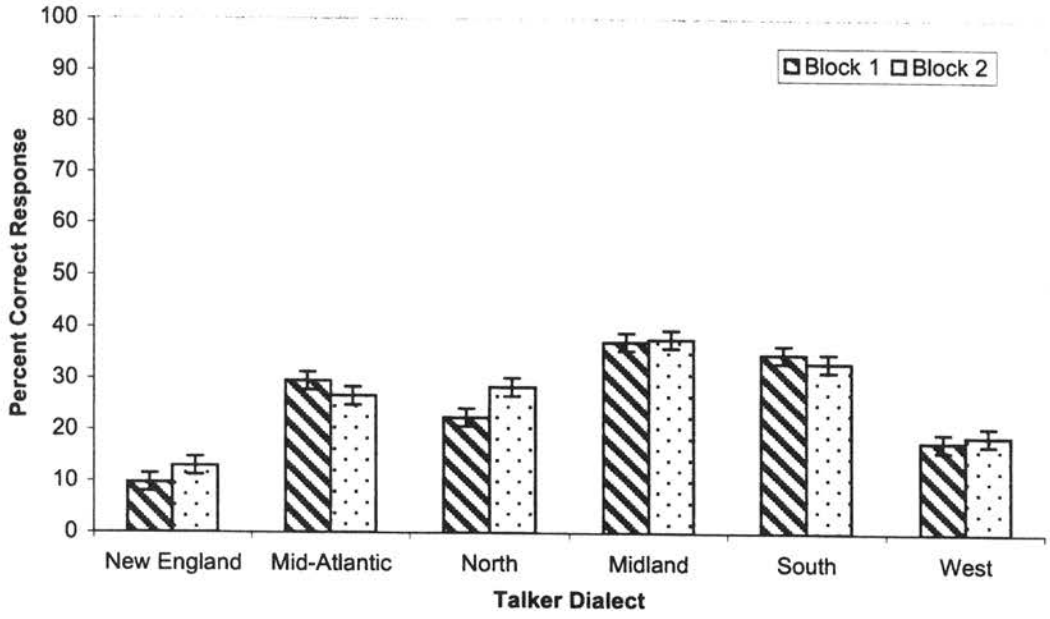


Figure 3. Percent correct categorization for each of the six talker dialect groups in each of the two experimental blocks, collapsed across listener group. Error bars indicate standard error.

Table 3 shows the performance of each of the four listener groups on each of the six talker dialects. As confirmed by the repeated measures ANOVA, performance is highly consistent across the four listener groups with respect to talker dialect.

Table 3. Percent correct categorization performance for each listener group for each talker dialect, collapsed across experimental block.

Listener Group	New England	Mid-Atlantic	North	Midland	South	West	Mean
Mobile North	10	26	27	39	39	18	27
Mobile Midland	12	29	26	40	33	16	26
Non-Mobile North	10	29	26	36	34	21	26
Non-Mobile Midland	13	29	23	35	31	19	25
Mean	11	28	26	38	34	19	26

Stimulus Repetitions

The mean number of stimulus repetitions for each of the listener groups, collapsed across experimental block, is shown in the last column of Table 2. A repeated measures ANOVA on stimulus repetitions with experimental block (first or second) as a within-subject variable and listener group (Mobile North, Mobile Midland, Non-Mobile North, or Non-Mobile Midland) as a between-subject variable revealed a significant main effect of experimental block ($F(1, 95) = 6.6$, $p = .012$). The

listeners repeated the stimulus items more often in the first block of trials (M = 1.74) than in the second block of trials (M = 1.65). Neither the main effect of listener group nor the block by listener group interaction were significant. Thus, across all four listener groups and both experimental blocks, the participants chose to listen to the stimulus items an average of 1.69 times before making their response and they repeated fewer stimuli as the experiment progressed from the first block to the second block of trials.

Perceptual Similarity

Previous research on dialect categorization performance has shown that residential history can affect the perceptual similarity space of dialect categories of naïve listeners (Clopper & Pisoni, 2004a, b). Therefore, in addition to examining the listeners' performance in terms of overall categorization accuracy, we also explored the patterns of errors produced by each of the four listener groups. This perceptual similarity analysis involved two phases. First, similarity and bias parameters were extracted from the raw categorization confusion data using Luce's (1963) and Shepard's (1957) Similarity Choice Model (SCM). Second, the resulting similarity parameters were submitted to an additive clustering analysis (Corter, 1982; Sattath & Tversky, 1977) to produce a graphical model of the perceptual similarity of the six regional dialects.

For each listener group, a 6 x 6 stimulus-response confusion matrix was calculated based on the listeners' responses in the six-alternative forced-choice task. The confusion matrices were summed over the two experimental blocks and the stimulus items were grouped by dialect. The four listener group confusion matrices were submitted to the full Similarity Choice Model (Luce, 1964; Shepard, 1957) to determine similarity and bias parameters. The similarity parameters indicate the degree of similarity between each pair of dialects. Similarity is assumed to be symmetric, so that the similarity between two categories i and j is equivalent to the similarity between j and i. The bias parameters provide an index of the response biases of the listeners and are discussed in more detail in the next section. Taken together, the similarity and bias parameters produce a model of perceptual categorization data which reveals the perceptual similarity of a set of objects or concepts (e.g., Nosofsky, 1985; Smith, 1980). In the current experiment, the objects were the six dialect regions: New England, Mid-Atlantic, North, Midland, South, and West.

A restricted version of the SCM analysis was also conducted in which the similarity parameters were held constant across all four listener groups, while the bias parameters were free to vary. If the restricted model produced results that did not fit the original data significantly worse than the full model, in which the similarity parameters were also free to vary across listener groups, then we would have evidence that the structure of the perceptual similarity spaces of the regional dialects of American English was equivalent for the four listener groups. In fact, however, the restricted SCM analysis produced a significantly worse fit to the data than the full model. The listener groups therefore differed in their perceptual similarity spaces of dialect variation.

To assess the effects of the independent variables of mobility and location on perceptual similarity, we also compared the perceptual similarity spaces of pairs of listener groups, using the restricted SCM. Significant differences were found in similarity structure between the Mobile North and Mobile Midland groups, the Mobile North and Non-Mobile North groups, and the Non-Mobile North and Non-Mobile Midland groups. However, the Non-Mobile Midland and Mobile Midland groups were not significantly different (G^2_{test} = 6.0, df = 15, χ^2_{crit} = 7.3, $p > .05$). Thus, the perceptual similarity spaces of the two Midland listener groups were equivalent, whereas the two Northern listener groups differed from each other and from their Midland counterparts.

The similarity parameters that were obtained from the SCM analyses were submitted to ADDTREE, an additive clustering scheme that produced graphical representations of perceptual similarity in tree form (Corter, 1982). For the Mobile North and Non-Mobile North groups, the similarity parameters from the full SCM analyses were used. For the Midland groups, the similarity parameters from the restricted model were used and a single tree representation was produced, because the restricted SCM analysis revealed equivalent perceptual similarity spaces across the two Midland listener groups.

The results of the ADDTREE analysis are shown in Figure 4. In these representations, perceptual similarity is inversely related to vertical distance. That is, the perceptual dissimilarity of any two dialect regions is represented by the sum of the lengths of the least number of vertical branches required to connect them. Horizontal distances are irrelevant in the interpretation of the figures.

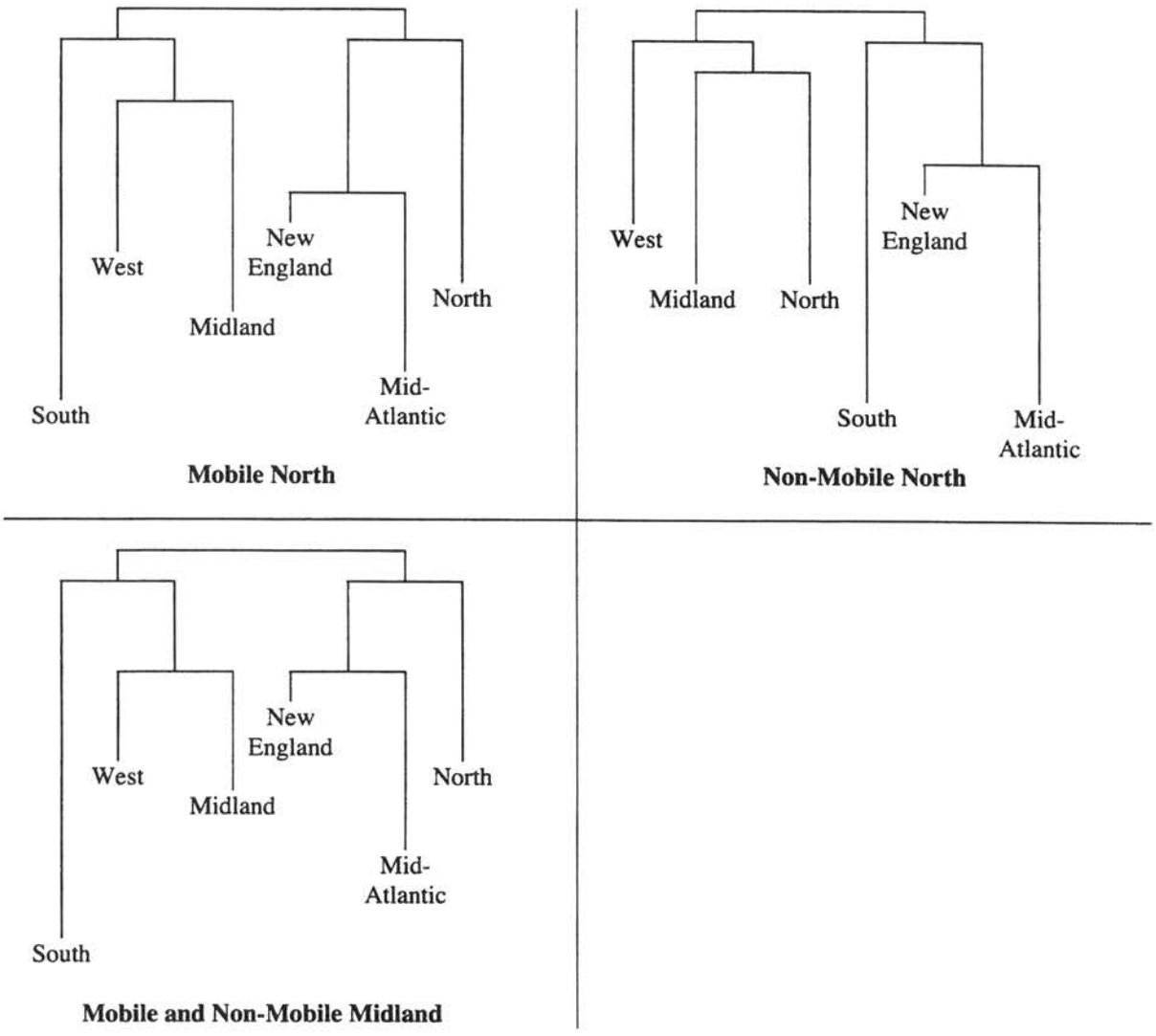


Figure 4. Clustering solutions for the Mobile North, Non-Mobile North, and Midland listeners.

A general pattern of perceptual similarity is evident across all three listener groups (Mobile North, Non-Mobile North, and Midland). First, in all three panels of Figure 4, Southern and Mid-Atlantic talkers are farthest from the root, suggesting that they are perceptually the most distinctive dialects of those examined in the current study. Second, the New England and Mid-Atlantic talkers cluster together perceptually for all three groups, although New England is always less distinct from the other dialects than the Mid-Atlantic. Finally, the Western and Midland talkers are also perceptually similar across all three listener groups. Thus, all of the listener groups appear to have categories for Southern talkers, Northeastern (New England and Mid-Atlantic) talkers, and unmarked (Midland and Western) talkers.

The perception of Northern talkers appears to be affected by the residential history of the listeners. In particular, while the structure of the trees for the Midland and Mobile Northern listener groups are virtually identical, the structure of the similarity model for the Non-Mobile Northern group is different. Specifically, the Northern talkers are relatively closely linked to the Mid-Atlantic and New England talkers in the trees for the Midland and Mobile Northern listeners. For the Non-Mobile Northern listeners, however, the Northern talkers are linked more closely to the Midland talkers. These differences suggest that the Non-Mobile Northern listeners perceived the Northern talkers as being more similar to the Midland talkers than the other groups did. That is, the listeners who had only lived in the Northern dialect region perceived themselves as being quite similar to their Midland peers. On the other hand, the listeners who had lived in the Midland dialect region for their whole lives and the listeners who had lived in more than one dialect region perceived the Northern talkers as being less similar to the Midland talkers and more similar to the New England and Mid-Atlantic talkers.

The highly similar perceptual structures revealed by the ADDTREE analysis for the Midland and Mobile Northern listener groups is striking. However, a Similarity Choice Model analysis of the stimulus-response confusion matrices restricting similarity parameters across the two Midland listener groups (Mobile and Non-Mobile) and the Mobile North listener group provided a significantly worse fit than the unrestricted full model, suggesting significant underlying differences between the Midland and Mobile Northern listeners. A closer inspection of the similarity parameters produced by the SCM analyses and the tree models produced in the ADDTREE analysis revealed that the Mobile Northern listeners were better able to distinguish the Southern and Mid-Atlantic talkers from the other talker dialect groups than the Midland listeners. In Figure 4, this greater perceptual distinctiveness is seen in the relatively longer lengths of the branches for the Northern, Midland, and Western talkers for the Mobile Northern listeners than for the Midland listeners. This finding suggests that although the Southern and Mid-Atlantic talkers were the most distinctive across all three listener groups, this perceptual distinctiveness was greater for the Mobile Northern listeners than for the Midland listeners. This difference is reflected in the somewhat better overall categorization performance by the Mobile Northern listeners than the Midland listeners (see Table 2).

Response Biases and Asymmetries

The Similarity Choice Model analysis (Luce, 1963; Shepard, 1957) also produced response bias parameters in addition to the similarity parameters used in the additive clustering analysis. The bias parameters provide an indication of the response biases of the listeners. In a forced-choice categorization task in which each category is presented an equal number of times in the stimulus materials, we would predict that the response biases of the listeners would be roughly equivalent across all of the response alternatives. In particular, with six categories, each response alternative would be selected 1/6 or 17% of the time, if the listeners were unbiased in their responses.

The observed response biases of the four groups of listeners are shown in Table 4. The bias parameters for the Northern listener groups (Mobile and Non-Mobile) are based on the full SCM analysis. The bias parameters for the Midland listener groups (Mobile and Non-Mobile) are based on the restricted model in which similarity parameters were held constant across both groups, because the restricted SCM analysis revealed equivalent similarity parameters for the two Midland groups. The bias parameters reveal a tendency across all four listener groups to display a positive bias towards Midland responses and a negative bias towards New England responses. That is, the listeners responded "Midland" more often and "New England" less often than if they were theoretically unbiased in their responses. The bias parameters for the other four talker groups are fairly close to the unbiased .17 response rate, suggesting little response bias for those categories.

Table 4. Bias parameters produced in the Similarity Choice Model analysis for each of the four listener groups. The bias parameters for the Northern listener groups (Mobile and Non-Mobile) are based on the full SCM analysis. The bias parameters for the Midland listener groups (Mobile and Non-Mobile) are based on the restricted model in which similarity parameters were held constant across both groups. Bias parameters close to .17 indicate relatively unbiased responses.

Listener Group	New England	Mid-Atlantic	North	Midland	South	West
Mobile North	.09	.14	.18	.28	.17	.14
Mobile Midland	.09	.17	.18	.28	.15	.13
Non-Mobile North	.09	.14	.18	.26	.16	.17
Non-Mobile Midland	.10	.17	.15	.27	.15	.15
Mean	.09	.16	.17	.27	.16	.14

One additional aspect of the participants' stimulus-response patterns was captured by the Similarity Choice Model analysis: asymmetrical response patterns. Asymmetrical patterns of similarity are not uncommon in categorization or similarity ratings experiments. In one classic example, participants will typically rate the similarity of North Korea to China as being greater than the similarity of China to North Korea, because China is perceived as being an appropriate baseline for comparison whereas North Korea is not (Tversky, 1977). In the present study, the raw confusion data produced by the naïve participants in the six-alternative forced-choice task suggest that this type of perceptual asymmetry may also be the present for the categorization of regional varieties of American English.

Table 5 shows the stimulus-response confusion matrix collapsed across all 99 listeners who participated in the six-alternative categorization task. Stimuli are presented in the rows and responses in the columns. One example of a perceptual asymmetry that was captured by the Similarity Choice Model analysis is the strong negative response bias for New England and the strong positive response bias for Midland. Table 5 shows that the New England talkers were categorized as North and Midland much more frequently than the North and Midland talkers were categorized as New England. However, in general, the New England response category was selected much less frequently than either the North or Midland response category and the bias parameters in Table 4 reflect this asymmetry.

Table 5. Mean proportion stimulus-response confusion matrix in the six-alternative forced-choice categorization task, collapsed across all 99 listeners.

		Response					
		New England	Mid-Atlantic	North	Midland	South	West
Stimulus	New England	.11	.15	.20	.28	.09	.17
	Mid-Atlantic	.25	.28	.13	.15	.08	.11
	North	.11	.13	.25	.28	.09	.14
	Midland	.07	.10	.17	.38	.07	.21
	South	.09	.09	.10	.23	.34	.14
	West	.08	.11	.20	.34	.09	.18

Another striking example of a perceptual asymmetry in similarity is the difference between the proportion of Mid-Atlantic talkers categorized as New England (.25) and the proportion of New England talkers categorized as Mid-Atlantic (.15). While the raw proportions are not too different, many more errors were made on the New England talkers than the Mid-Atlantic talkers overall. Thus, 35% of the incorrect responses to Mid-Atlantic talkers were New England responses, whereas only 17% of the incorrect responses for the New England talkers were Mid-Atlantic responses. The SCM and ADDTREE analyses revealed a high degree of similarity between the New England and Mid-Atlantic regions, however, masking this important asymmetry. The SCM analysis was unable to model this particular perceptual asymmetry due to the overall tendency for a low response bias for New England. Thus, while some of the stimulus-response asymmetries were reflected in the similarity and bias parameters produced in the SCM analyses, the fact that the Mid-Atlantic talkers were more often confused with New England talkers than vice versa was not.

Discussion

The results of the current study are consistent with previous research on the perceptual categorization of dialect variation in the United States and Europe. Overall performance as measured by accuracy was 26% correct, which is poor but statistically above chance. In addition, this level of performance is somewhat lower than the 31% accuracy reported by Clopper and colleagues for similar tasks (Clopper et al., in press; Clopper & Pisoni, 2004a, b). One important difference between our previous work and the present study, however, is the corpus from which the stimulus materials were obtained for presentation to the listeners. The earlier research was conducted using speech samples obtained from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher et al., 1986). As discussed above, the original design of the TIMIT corpus was less informed with respect to regional linguistic variation in the United States. First, the regional labels provided for each talker do not correspond to dialect regions based on current sociolinguistic research. Second, the criteria that were used to assign the regional label to each talker did not explicitly control for the residential history of the talkers and their parents.

The speech materials in the current study, however, were taken from the Nationwide Speech Project corpus (Chapter 2), which was designed specifically for perceptual and acoustic analyses of dialect variation in the United States. Therefore, the sociolinguistic components of the corpus were more carefully controlled and documented. The dialect labels and geographic regions included in each dialect were based on the most current sociolinguistic research by Labov and his colleagues (Labov et al., forthcoming). Details about the residential history of the talkers and their parents were obtained from each talker and only lifetime residents of each dialect region were included.

Based on the differences between the two corpora, we might expect that performance would be better in the present study than in the previous study because the talkers more accurately reflect regional dialect variation in the United States. One crucial difference, however, is the inclusion of talkers from both New England and the Mid-Atlantic in the current experiment. Although the TIMIT corpus contains talkers from New York City, they were too few in number to be included in the previous research (Clopper et al., in press; Clopper & Pisoni, 2004a, b). Therefore, overall performance in the current study may be lower overall because the listeners were forced to distinguish between New England and Mid-Atlantic talkers. The results of the accuracy and response asymmetry analyses both suggest that this particular aspect of the categorization task was quite difficult for the naïve listeners. In particular, the listeners in this study were quite poor in accurately categorizing New England talkers. In addition, they categorized Mid-Atlantic talkers as New Englanders about 25% of the time. Thus, the lower performance found in the present study may be a better reflection of the dialect categorization abilities of naïve listeners in the United States than the previous results reported by Clopper and her colleagues (Clopper et al., in press; Clopper & Pisoni, 2004a, b), which did not require the listeners to make distinctions between the two Northeastern varieties of American English, New England and Mid-Atlantic.

Despite the somewhat lower overall level of categorization performance, the results of the perceptual similarity analyses are consistent with previous research on the perceptual similarity of regional varieties of American English (Clopper et al., in press; Clopper & Pisoni, 2004b). Across the four listener groups, a common pattern of perceptual similarity emerged in which the South and the Mid-Atlantic were the most distinctive dialects. In addition, the New England talkers clustered with the Mid-Atlantic talkers to create a perceptual Northeastern dialect. The Midland and Western talkers also clustered together in a third salient dialect. Thus, the results of this six-alternative forced-choice task suggest that the three main perceptual dialects of American English for naïve listeners are Northeast, South, and Midwest/West. These three clusters are quite similar to the perceptual clusters described by Clopper and colleagues (Clopper et al., in press; Clopper & Pisoni, 2004a, b) based on our research using the TIMIT corpus.

The perceptual similarity spaces of the listeners are also consistent with the phonological properties of the dialects, as discussed in the sociolinguistics literature. For example, Labov (1998) described the three major dialects of American English as North, South, and the "Third Dialect" (which includes the West and the Midland), with the Mid-Atlantic as an exception. Carver (1987) defined the major dialects as North and South, with the Western region included in the North. Even as early as 1925, Krapp divided the varieties of American English into Eastern, Southern, and Western (or General) American groups.

For the specific set of talkers used in the current study, significant acoustic-phonetic differences were found for the Northern and Southern talkers, while the overall similarity between the Midland and Western talkers was quite high (Chapter 2). Thus, the perceptual similarity of the dialects reflects the overall phonological properties of the talkers in the NSP corpus. The phonological salience of the Mid-Atlantic talkers was not captured by the acoustic-phonetic analysis of the vowel spaces of the talkers in the NSP corpus, however. Further research on vowel, consonant, and prosodic variation in the speech of the talkers in the NSP corpus will be needed to determine which acoustic-phonetic properties contributed to the perceptual distinctiveness of the Mid-Atlantic talkers (see Docherty & Foulkes, 1999).

The perceptual similarity of the Northern talkers with respect to the other dialects was affected by the residential history of the listeners. In particular, the Non-Mobile Northern listeners perceived the Northern talkers as being most similar to the Midland talkers, whereas the Midland and Mobile Northern listeners perceived the Northern talkers as being more similar to the New England talkers.

This result is interesting for several reasons. First, it confirms that a listener's residential history affects the perception of dialect variation. For the Northern listeners, mobility was a contributing factor in the perceptual similarity of the Northern and Midland dialects. At the same time, for the Non-Mobile listeners, location was relevant to perceptual similarity of these same two dialects.

Second, the fact that the lifetime residents of the Northern dialect region (the Non-Mobile Northern listener group) did not attend to the difference between Northern and Midland talkers is consistent with a study by Niedzielski (1999) that examined the perception of the Northern Cities Chain Shift. In her study, Niedzielski (1999) asked listeners in Detroit to match natural vowel stimuli to synthetically produced vowel tokens based on vowel quality. The listeners were presented with sentence-length utterances read by a female talker from the Detroit area and were asked to pay attention to a target word in the sentence. They were then presented with six synthetic vowel stimuli that included a range of first and second formant frequencies and were asked to select the vowel token that was the best match to the target. Prior to the beginning of the experiment, half of the listeners were told that the talker was from Detroit and the other half were told that the talker was from Canada. Niedzielski (1999) found that the label she provided about the talker's region of origin had a significant effect on the listeners' performance. In particular, the listeners in the Detroit group consistently selected canonical, unshifted vowels as the best match, whereas the listeners in the Canadian group selected vowel tokens that more closely matched the talker's actual productions. The results of Niedzielski's (1999) study suggest that Northern listeners may not perceive the vowel shifts that are present in their own speech. Similarly, in the present study, the fact that the Non-Mobile Northern listener group perceived the Northern and Midland talkers as being highly similar, despite robust acoustic-phonetic differences between the two groups, suggests that listeners who have lived only in the Northern dialect region may not perceive the phonological differences between themselves and Midland talkers.

On the other hand, participants who were not from the North and those participants who had lived in multiple dialect regions perceived the Northern and New England talkers as being more similar than the Northern and Midland talkers. This similarity is supported by the acoustic analysis of the talkers described in Chapter 2 and by other research suggesting that New England may be the geographic origin of the Northern Cities Chain Shift, resulting in some phonological similarities between New England and Northern talkers (Boberg, 2001). Thus, the perceptual similarity spaces of the Midland and Mobile Northern listener groups may more accurately reflect the phonological similarities of the different regional dialects than the Non-Mobile Northern listener group.

The common aspects of the residential history of the listeners influenced the response biases that were observed consistently across all four listener groups. In particular, the negative bias towards responding "New England" may be due to a general unfamiliarity with New England speech. Of the Mobile listeners, only five out of 49 (10%) had lived in New England for any period of time. The listeners who had not lived in New England may have been particularly limited in their exposure to New England speech due to the relatively small number of students at Indiana University who come from the New England area. Table 6 shows the percentage of entering undergraduate students in 2002 from each of the six dialect regions included in the Nationwide Speech Project corpus. It is clear that even in a university setting, participants may not have encountered very many talkers from New England. This general unfamiliarity with New England speech may have led to the strong negative bias for "New England" responses.

Table 6. Percentage of 2002 Indiana University first year students from each of the six dialect regions included in the Nationwide Speech Project corpus.

Region	Percentage of Entering Students
New England	<1
Mid-Atlantic	4
North	45
Midland	42
South	4
West	2
Other	2

Similarly, the positive bias for the Midland talkers may have been due to the extreme familiarity of the listeners with talkers from this region. Indiana University is located in the Midland dialect region and the Midland dialect is also one of the best represented dialect regions at the University. The listeners may have adopted "Midland" as a default benchmark response for all talkers who sounded like themselves, resulting in a large positive bias for the "Midland" response.

Familiarity, or lack thereof, may also be related to the asymmetries found in the stimulus-response confusion matrices. In particular, the fact that Mid-Atlantic talkers were frequently identified as New Englanders but New Englanders were most often misidentified as Midlanders is consistent with the claim that the listeners in the current study were unfamiliar with the New England dialect. In addition, these results also suggest that the listeners recruited for participation in this study may have only a single category for the Northeastern dialects and that they expect talkers from the entire Northeastern region to sound like Mid-Atlantic talkers. That is, the listeners do not know which phonological properties are associated with New England talkers and which ones are unique to Mid-Atlantic talkers. An alternative interpretation is that the listeners were simply not attending to the distinction between the two Northeastern groups, like the Non-Mobile Northerners' apparent inattention to the differences between Northern and Midland talkers. Future research should examine the perceptual categorization performance of listeners from the Mid-Atlantic and New England regions to determine whether this effect is due to familiarity and exposure or to a general lack of perceptual attention to the phonological differences between these two varieties.

The listeners in the current study were all young adults who differed in their linguistic experience and exposure to dialect variation on two specific dimensions: geographic mobility and location. The results of the analyses of the underlying perceptual similarity spaces revealed that both of these factors contribute to the representations and categories that listeners form about the regional varieties of their native language. In particular, mobility can increase the distinctiveness of different varieties, presumably as a result of greater experience with specific varieties and language variation in general. Frequency of exposure has been shown to play an important role in the development of robust categories in the laboratory, particularly when the participants' experience involves not only a specific item, but also the appropriate category label for that item (Barsalou, 1985). By living in several different regions of the country, the mobile listeners had the opportunity to develop more reliable categories through greater exposure to talkers from different dialect regions *and* implicit knowledge of the appropriate category label for those talkers.

Location, on the other hand, can reduce the distinctiveness of certain varieties due to commonly held social beliefs or stereotypes. In the present study, the Non-Mobile Northerners perceived the Northern and Midland talkers as being more similar to each other than the other listeners did, presumably because they did not perceive the differences between the two dialects and they

believed that the two groups are highly similar in their speech. In terms of categorization, this finding suggests a shrinking of the perceptual distance between the Northern and Midland talkers for the Non-Mobile Northern listeners due to a lack of attention to the phonological differences between the two dialects (see Nosofsky, 1986). Additional research is needed to determine whether this perceptual shrinking is due to misperception of the linguistic properties of the signal or perceptual biases, such as those reported by Niedzielski (1999) in the vowel-matching task.

Conclusions

Naïve listeners categorized the talkers in the NSP corpus by regional dialect with 26% accuracy overall. While the residential history of the listeners did not affect the accuracy of their performance in the six-alternative forced-choice task, it did contribute to the patterns of confusions made by the listeners. In particular, while all four groups of listeners perceived three main dialect categories: Northeast, South, and Midwest/West, the Non-Mobile Northern listener group perceived a greater similarity between the Northern and Midland talkers than the Mobile Northern and Midland listeners did. These results are consistent with previous research showing the effects of the geographic mobility and location of the listeners on the perception of dialect variation. The relationship between linguistic experience and speech perception thus appears to extend beyond acquisition of phonological systems to the perception and categorization of phonologically different regional varieties. In particular, geographic mobility and location affect the perceptual similarity space of regional varieties of American English for naïve listeners, just as early linguistic experience with specific phonemic contrasts has lasting effects on the perception of those contrasts into adulthood.

References

Aslin, R.N., & Pisoni, D.B. (1980). Linguistic experience modifies lexical stress perception. *Journal of Child Language, 10*, 535-549.

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Boberg, C. (2001). The phonological status of Western New England. *American Speech, 76*, 3-29.

Bowie, D. (2000). *The Effect of Geographic Mobility on the Retention of a Local Dialect.* Doctoral dissertation, University of Pennsylvania, Philadelphia.

Bruer, J.T. (2001). A critical and sensitive period primer. In D.B. Bailey, Jr., J.T. Bruer, F.J. Symons, & J.W. Lichtman (Eds.), *Critical Thinking about Critical Periods* (pp. 3-26). Baltimore: Paul H. Brookes.

Carver, C.M. (1987). *American Regional Dialects: A Word Geography.* Ann Arbor, MI: University of Michigan Press.

Chambers, J.K. (1992). Dialect acquisition. *Language, 68*, 673-705.

Chambers, J.K. (2002). Dynamics of dialect convergence. *Journal of Sociolinguistics, 6*, 117-130.

Clopper, C.G., Conrey, B.L., & Pisoni, D.B. (in press). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology.*

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.

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

Docherty, G.J., & Foulkes, P. (1999). Derby and Newcastle: Instrumental phonetics and variationist studies. In P. Foulkes & G.J. Docherty (Eds.), *Urban Voices* (pp. 47-71). London: Edward Arnold.

Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society, 17*, 183-207.

Evans, B.G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115*, 352-361.

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

Flege, J.E., Munro, M.J., & MacKay, I.R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*, 3125-3134.

Hakuta, K. (2001). A critical period for second language acquisition? In D.B. Bailey, Jr., J.T. Bruer, F.J. Symons, & J.W. Lichtman (Eds.), *Critical Thinking about Critical Periods* (pp. 193-205). Baltimore: Paul H. Brookes.

Jusczyk, P.W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Krapp, G.P. (1925). *The English Language in America*. New York: Frederick Ungar.

Labov, W. (1971). Stages in the acquisition of standard English. In H.B. Allen & G.N. Underwood (Eds.), *Readings in American Dialectology* (pp. 473-499). New York: Appleton-Century-Crofts.

Labov, W. (1994). *Principles of Linguistic Change: Internal Factors*. Malden, MA: Blackwell.

Labov, W. (1998). The three dialects of English. In M.D. Linn (Ed.), *Handbook of Dialects and Language Variation* (pp. 39-81). San Diego: Academic Press.

Labov, W., Ash, S., & Boberg, C. (forthcoming). *Atlas of North American English*. New York: Mouton de Gruyter.

Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 103-189). New York: Wiley.

Markham, D. (1997). *Phonetic Imitation, Accent, and the Learner*. Lund, Sweden: Lund University Press.

Mason, H.M. (1946). Understandability of speech in noise as affected by region of origin of speaker and listener. *Speech Monographs, 13(2)*, 54-58.

McDavid, R.I., Jr. (1958). The dialects of American English. In W.N. Francis (Ed.), *The Structure of American English* (pp. 480-543). New York: Ronald Press.

Munro, M.J., Derwing, T.M., & Flege, J.E. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics, 27*, 385-403.

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

Nosofsky, R.M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice-model analysis. *Perception and Psychophysics, 38*, 415-432.

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Payne, A.C. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In W. Labov (Ed.), *Locating Language in Time and Space* (pp. 143-178). New York: Academic Press.

Plichta, B., & Preston, D. R. (2003). The /ay/s have it: Stereotype, perception, and region. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

Politzer, R.L. (1993). A researcher's reflections on bridging dialect and second language learning: Discussion of problems and solutions. In B.J. Merino, H.T. Trueba, & F.A. Samaniego (Eds.), *Language and Culture in Learning: Teaching Spanish to Native Speakers of Spanish* (pp. 45-57). Washington, DC: Falmer.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics, 52*, 37-52.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America, 97*, 1286-1296.

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.

Rakerd, B., & Plichta, B. (2003). More on perceptions of /ɑ/ fronting. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

Roberts, J. (1997). Hitting a moving target: Acquisition of sound change in progress by Philadelphia children. *Language Variation and Change, 9*, 249-266.

Roberts, J. (2002). Child language variation. In J.K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 333-348). Malden, MA: Blackwell.

Ryan, E.B., & Giles, H. (1982). *Attitudes towards Language Variation.* London: Edward Arnold.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*, 319-345.

Scovel, T. (1988). *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech.* Cambridge, MA: Newbury House.

Shepard, R.N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22*, 325-345.

Smith, J.E.K. (1980). Models of identification. In R. Nickerson (Ed.), *Attention and Performance VIII* (pp. 129-158). Hillsdale, NJ: Erlbaum.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 3-45). Timonium, MD: York Press.

Tees, R.C., & Werker, J.F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to 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.

Thomas, E.R. (2002). Sociophonetic applications of speech perception experiments. *American Speech, 77*, 115-147.

Tice, R., & Carrell, T. (1998). Level16 (Version 2.0.3) [Computer Software]. Lincoln, NE: University of Nebraska.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology, 18*, 31-48.

Van Bezooijen, R., & Ytsma, J. (1999). Accents of Dutch: Personality impression, divergence, and identifiability. *Belgian Journal of Linguistics, 13*, 105-129.

Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D.R. Preston (Ed.), *Handbook of Perceptual Dialectology* (pp. 345-358). Philadelphia: John Benjamins.

Yamada, R.A. (1995). Age and acquisition of second language speech sounds: Perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 305-320). Timonium, MD: York Press.

CHAPTER 4: FREE CLASSIFICATION OF DIALECTS

Introduction

Previous research on the perception of dialect variation has typically involved one of two main approaches. The first of these approaches is perceptual dialectology. The goal of this type of research is to explore naïve participants' beliefs and attitudes about regional and social variation (Preston, 1999). Preston (1989, 1993) has been the primary researcher in perceptual dialectology in the United States, but similar research has also been conducted in Japan, Great Britain, and Germany (see Preston, 1999). The second primary research approach is sociophonetics, which draws on methods developed in acoustic-phonetics and speech science to investigate the perception and cognitive representation of dialect variation (Thomas, 2002). Clopper and her colleagues (Clopper, Conrey, & Pisoni, in press; Clopper & Pisoni, 2004a, b) have conducted a series of perceptual categorization experiments to examine the explicit identification of talker dialect. Other researchers (e.g., Niedzielski, 1999; Plichta & Preston, 2003; Rakerd & Plichta, 2003) have explored the perception of specific linguistic variables as they relate to regional variation in the United States. Similar research has also been conducted in Great Britain (Evans & Iverson, 2004; Williams, Garrett, & Coupland, 1999) and the Netherlands (Van Bezooijen & Gooskens, 1999; Van Bezooijen & Ytsma, 1999).

Perceptual Dialectology

One of the primary methods used in perceptual dialectology is the map-drawing task. Participants are given a map of the United States, including state boundaries, and are asked to draw and label areas where people speak differently (Preston, 1993). Naïve participants varied greatly in their performance of this task, indicating as few as two regions (North and South) or as many as 15 or more. Participants also used a large range of labels to describe the regions that they drew; Preston (1986) compiled a list of 25 different geographic terms from the individual maps drawn by his participants. Despite these individual differences, however, composite data across all of the participants revealed eight to 12 dialect regions, including salient cultural and linguistic regions such as the South, New York City, and the Midwest (Preston, 1986).

More recently, Tamasi (2003) explored naïve participants' beliefs about regional variation without constraining the dialect categories geographically. Using a free classification task (Imai, 1966; Imai & Garner, 1965), she gave participants a stack of index cards with the names of the 50 states and asked them to group the cards by how people talked in each state. As in the map-drawing task, participants varied in their classification strategies, creating as few as five groups and as many as 35, with a mean of 14. Composite data, however, revealed six or seven dialect regions, with some of the most salient regions corresponding to the South, New England, and the Midwest.

While the map-drawing task used by Preston (1986) and the free classification task with index cards used by Tamasi (2003) provide some information about the beliefs of naïve participants with respect to regional dialect variation in the United States, a number of issues remain unresolved. First, geographic knowledge of the United States, or lack thereof, may play an important role in the performance of the participants in these two studies. Preston (1986) admitted that it was necessary to provide the participants with a map that included state boundaries due to the limited geographic knowledge of most Americans. While Tamasi (2003) tried to avoid issues related to geography by using the free classification method, some of the results that she discussed suggest that a lack of geographic knowledge may have been a contributing factor in some of her participants' responses. For example, she described one participant who put Texas, Oklahoma, and Wyoming in a group together because those are the states in which cowboys live and, therefore, "cowboy speech" is represented in

each state. Tamasi (2003) assumed that the participant was aware of the geographic distance between Wyoming and the other two states and made the classification based on some other factor. However, it is also possible that the participant believed that all three states were geographically contiguous.

Second, in these kinds of experiments, the participants were not actually presented with samples of speech, but instead had to rely on their memory or stereotypes of regional variation in making their responses. It is therefore possible that the maps or similarity groups they produced were based on general cultural categories and not on the perception of linguistic differences. Preston (1986) made a similar observation, pointing out that the maps drawn by naïve participants did not correspond to maps of dialect variation produced by sociolinguists based on actual linguistic variation, but that the maps drawn by his participants instead resembled culturally-based maps of the United States. For example, the Northern and Midland dialect regions are typically collapsed into a single Midwest cultural region in the maps drawn by naïve participants and other salient cultural regions such as the Northeast and the South appear repeatedly in the perceptual dialectology studies (Ayers, Limerick, Nissenbaum, & Onuf, 1996).

Dialect Categorization

Several recent studies on the perception of dialect variation have used speech samples of talkers from different regions of the United States to examine naïve listeners' identification and categorization of different linguistic varieties. Clopper and Pisoni (2004b) asked naïve listeners to categorize unfamiliar male talkers by regional dialect using sentence-length materials in a six-alternative forced-choice categorization task. The stimulus materials were taken from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986) and the talkers represented six regional varieties of American English: New England, North, North Midland, South Midland, South, and West. While overall accuracy was only 31%, categorization performance was statistically above chance. In addition, the large number of errors produced by the listeners provided sufficient data for an analysis of the perceptual similarity of the dialects. Using the stimulus-response confusion matrices from the six-alternative task, we conducted a clustering analysis to explore the listeners' perceptual dialect categories. The results of this analysis revealed three main dialect categories: New England, South, and Midwest/West. Clopper et al. (in press) replicated these findings with a set of female talkers and a mixed set of male and female talkers. As in the earlier study, the patterns of errors revealed reliable perceptual dialect categories corresponding to New England, the South, and the Midwest/West.

Similar categorization results have also been obtained using stimulus materials from the Nationwide Speech Project (NSP) corpus (Chapter 2). In Chapter 3, naïve listeners were presented with meaningful English sentences spoken by a set of male and female talkers from six different regions of the United States: New England, Mid-Atlantic, North, Midland, South, and West. As in the earlier studies, performance was statistically above chance, although overall accuracy was only 26%. The stimulus-response confusion matrices were submitted to a clustering analysis to uncover the perceptual similarity of the dialects for the naïve listeners. The clustering analysis revealed three main dialect categories: Northeast, South, and Midwest/West. These categories are almost identical to the clusters found in the previous research using the TIMIT corpus (Clopper et al., in press; Clopper & Pisoni, 2004b). The minor differences between them can be attributed to the differences between the two sets of stimuli in terms of the dialect regions that were included.

One of the primary limitations of this set of studies is that the dialect regions and response labels were provided to the listeners by the experimenter. The perceptual dialectology research discussed above suggests that naïve participants may have a large range of labels and regions that they

consider to be culturally and linguistically salient. In order to reduce the constraints imposed by verbal labels for the listeners, the current study explored the perceptual structure of regional varieties of American English using a free classification procedure. This type of perceptual task allows the participants to create groups based on their own perceptual categories, without imposing an a priori geographic structure on those categories.

Free classification paradigms were developed in cognitive psychology to study the perceptual dimensions of stimuli (Imai, 1966; Imai & Garner, 1965). The unconstrained nature of the task allows experimenters to explore the primary stimulus dimensions that naïve participants attend to in making their classifications. In some versions of the free classification task, participants are asked to sort stimuli into a specific number of groups, such as two or four. The results of this type of free classification task provide insight into the most salient or important dimensions of classification (Imai & Garner, 1965; Medin, Wattenmaker, & Hampson, 1987). In other free classification experiments, participants are permitted to make as many groups as they want and the aggregate data are submitted to scaling analyses, such as multidimensional scaling, to extract underlying perceptual similarity dimensions (McAdams, 1993). The free classification experiments conducted in the present study were designed to investigate the underlying dimensions of perceptual similarity for regional varieties of American English and the task was not constrained to require the participants to make a specified number of groups.

Linguistic Experience and the Perception of Dialect Variation

One factor that has repeatedly been shown to affect the perception of dialect variation is the linguistic experience of the participants. Preston (1986) reported the results of his map-drawing task from participants in Hawaii, Michigan, Indiana, upstate New York, and New York City. While the overall results were consistent across all of the listener groups, he did find some differences between the groups due to their different locations in the United States. First, he found that the participants produced more finely graded regions for areas that were geographically close to home compared to areas farther away. For example, the Indiana participants divided the Midwest area into three distinct regions (North, Midwest, and Indiana). Meanwhile, the participants from New York City divided the Northeast into five different regions (New York City, Boston, New England, North, and New York State). These results suggest that personal experience affects naïve participants' beliefs about dialect variation in speech. In particular, exposure to local variation led to more perceived differences and finer categories close to home, whereas a lack of exposure to variation in other parts of the country led to broader categories in those regions. These findings are consistent with research on categorization behavior that demonstrates important effects of stimulus frequency. Naïve participants develop more fine-grained categories for local linguistic varieties than for varieties with which they have less experience because they encounter more people from local areas than from more geographically distant areas (Murphy, 2002; Nosofsky, 1988).

Tamasi's (2003) free classification experiment was conducted with participants from Georgia and New Jersey. As in the map-drawing task, the overall pattern of results was similar across the two groups of participants. However, the New Jersey participants made a clear distinction between New England and the Mid-Atlantic which the participants from Georgia did not. Meanwhile, the participants from Georgia made a sharper distinction between the Southeast and the Gulf States than the New Jersey participants did. Again, these results suggest that the participants' geographic location and their experience with local and neighboring dialects affected their beliefs about linguistic variation in the United States.

The role of linguistic experience has also been explored in dialect categorization performance (Chapter 3; Clopper & Pisoni, 2004a). In one recent study, we recruited two groups of listeners to participate in a six-alternative forced-choice dialect categorization task using the TIMIT stimulus materials (Clopper & Pisoni, 2004a). The listeners in one group, the "mobile" group, had lived in at least three different states at the time of testing. The second group of listeners, the "non-mobile" group, had all lived only in Indiana at the time of testing. A comparison of the categorization performance by these two groups revealed that the mobile listeners were slightly more accurate than the non-mobile listeners in categorizing the unfamiliar talkers by regional dialect. In addition, a post-hoc analysis comparing listeners who had lived in each dialect region to listeners who had not lived there revealed a significant difference in performance for residents and non-residents. Specifically, the residents of each region performed better than the non-residents in categorizing talkers from that dialect region, suggesting that direct exposure to different dialects as a result of residential history affects categorization accuracy.

A clustering analysis of the stimulus-response confusion matrices for each listener group revealed the same three clusters reported by Clopper and Pisoni (2004b): New England, South, and Midwest/West. However, some significant differences were found between the listener groups with respect to the underlying structure of the perceptual similarity spaces. In particular, the mobile listeners tended to perceive greater differences between geographically contiguous regions, such as the North and North Midland or the North Midland and South Midland, than the non-mobile listeners did.

In Chapter 3, we also found significant differences in perceptual similarity of the dialects due to the linguistic experience of the listeners. Four groups of listeners participated in a six-alternative forced-choice dialect categorization task using the NSP stimulus materials. The listener groups represented two variables related to residential history: geographic mobility (Mobile and Non-Mobile) and location (North and Midland). While all four groups performed the task with the same degree of accuracy, a clustering analysis on the stimulus-response confusion matrices revealed significant effects of both geographic mobility and location. First, the Mobile and Non-Mobile Northerners differed in their perception of the Northern talkers, with the Non-Mobile Northerners perceiving them as less distinctive from the other dialects than the Mobile Northern listeners. Second, the overall perceptual similarity structure of the dialects for the Mobile North and the Midland listeners were identical, but the Mobile North listeners perceived greater distinctiveness between the Southern and Mid-Atlantic talkers and the Midland and Western talkers than the Midland listeners did.

All four of these studies showed an effect of the location of the listeners, with a tendency for greater perceptual distinctiveness and categorization accuracy for local varieties. The one exception to this trend was the Northern listeners, who were found to misperceive their own dialect (Chapter 3). The two studies by Clopper and colleagues (Chapter 3; Clopper & Pisoni, 2004a) also revealed evidence of an effect of geographic mobility, with mobile listeners performing more accurately and perceiving greater distinctiveness between the different dialects.

The present set of experiments was designed to explore the role of residential history, specifically geographic mobility and location, using an unconstrained free classification task (Imai & Garner, 1965). Experiment 1 was a pilot study of the free classification paradigm using the TIMIT stimulus materials. The TIMIT materials were used in the pilot study to allow comparison between the results of the free classification task and the earlier research on dialect perception (Clopper & Pisoni, 2004a, b). The listeners in Experiment 1 were mixed with respect to their residential history. Experiment 2 used the free classification method developed in Experiment 1 with stimulus materials from the NSP corpus. In addition, four groups of listeners with different residential histories were recruited for participation in Experiment 2. The results of Experiment 2 were therefore directly

comparable to the results obtained from the six-alternative forced-choice categorization task using the NSP materials (Chapter 3).

Based on the results of the earlier forced-choice categorization studies (Chapter 3; Clopper & Pisoni, 2004a, b), we had several predictions regarding the outcome of the free classification experiments. First, given that naïve listeners appear to have three perceptual dialect categories, we predicted that they would create a relatively small number of groups of talkers in the free classification task. Second, we expected geographic mobility and location to affect performance, particularly with respect to the perception of a high degree of similarity between the Northern and Midland talkers for the Non-Mobile Northern listeners. Finally, the Mobile listeners were predicted to more accurately distinguish between talkers of different dialects and to create more groups of talkers as a result of more well-defined perceptual dialect categories.

Experiment 1

Methods

Listeners. Twenty-five listeners were recruited from the Indiana University community for participation in this experiment. Prior to the data analysis, data from three participants were removed: two had a history of a hearing or speech disorder reported at the time of testing and one was substantially older than any of the other participants. The remaining 22 listeners were between the ages of 18 and 25 years old. They were all monolingual native speakers of American English with no reported history of a hearing or speech disorder. Both parents of each participant were also native speakers of English. The participants received $8 for their service.

Talkers. Sixty-six male talkers from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher et al., 1986) produced the stimulus materials for Experiment 1. The 66 talkers were between the ages of 20 and 29 at the time of recording, with eleven talkers from each of six dialect regions in the United States: New England, North, North Midland, South Midland, South, and West.[12] The set of talkers used in the current experiment was identical to the set used in the previous forced-choice dialect categorization experiments by Clopper and Pisoni (2004a, b).

The six dialects represented by the talkers in Experiment 1 differed in terms of a number of segmental properties. Clopper and Pisoni (2004b) conducted an acoustic-phonetic analysis of the speech of the 66 talkers used in the current experiment and found a number of phonetic variables that distinguished the six different dialects. The New England talkers were r-less and exhibited [æ] backing in the word *rag*. The Northern talkers produced centralized [ow] offglides in the word *don't* and monophthongal [æ]s in the word *rag*. The South Midland dialect was characterized by [u] fronting in the word *suit* and [ow] backing in the word *don't*. The Southern dialect included the *greasy ~ greazy* alternation. The North Midland and Western talkers were not found to produce any marked dialect features, among those analyzed.

[12] The dialect region labels used in Experiment 1 refer to the labels provided with the original TIMIT corpus. While these dialect regions do not correspond to the dialect regions in the NSP corpus or to current sociolinguistic descriptions of regional variation in the United States, additional information about the residential history of the TIMIT talkers is no longer available, so reassignment of the talkers to more appropriate dialect groups was not possible.

Stimulus Materials. The stimulus materials consisted of one novel sentence per talker, for a total of 66 different sentences. Examples of the novel sentences from the TIMIT corpus are shown in (1). A list of the stimulus materials used in Experiment 1 is provided in Appendix D.

(1) Barb's gold bracelet was a graduation present.
A huge tapestry hung in her hallway.
Please shorten this skirt for Joyce.

The stimuli presented to the participants were 66 QuickTime movies created using FinalCut Pro. The auditory track of each movie contained one of the novel sentences. The original TIMIT .wav files had been edited to include only the speech material and the mean amplitude of each sentence was leveled to 55 dB using Level16 (Tice & Carrell, 1998). These edited .wav files served as the auditory track for the QuickTime movies. The movies differed in their overall duration, with the duration of each movie corresponding to the duration of the sound file contained in its auditory track. The visual track of each movie was a 765 x 765 pixel solid blue square outlined with a medium gray box that was approximately 35 pixels wide, for a total visual stimulus that was 800 x 800 pixels in size. Each movie was then scaled to 40 x 40 pixels at the time of rendering and sampled at a rate of 15 frames per second.

Procedure. The participants were seated at personal computers equipped with a two-button computer mouse and a set of Beyerdynamic DT100 headphones. On the computer screen, the participants saw 66 blue squares arranged in columns on the left and a 20 x 20 cell grid on the right. A schematic of the display is shown in Figure 1. The participants could listen to each movie by clicking on the blue square with the right mouse button. The auditory track of each movie was presented at approximately 70 dB SPL over the headphones. The participants could move the movie around the screen by clicking and dragging it with the left mouse button.

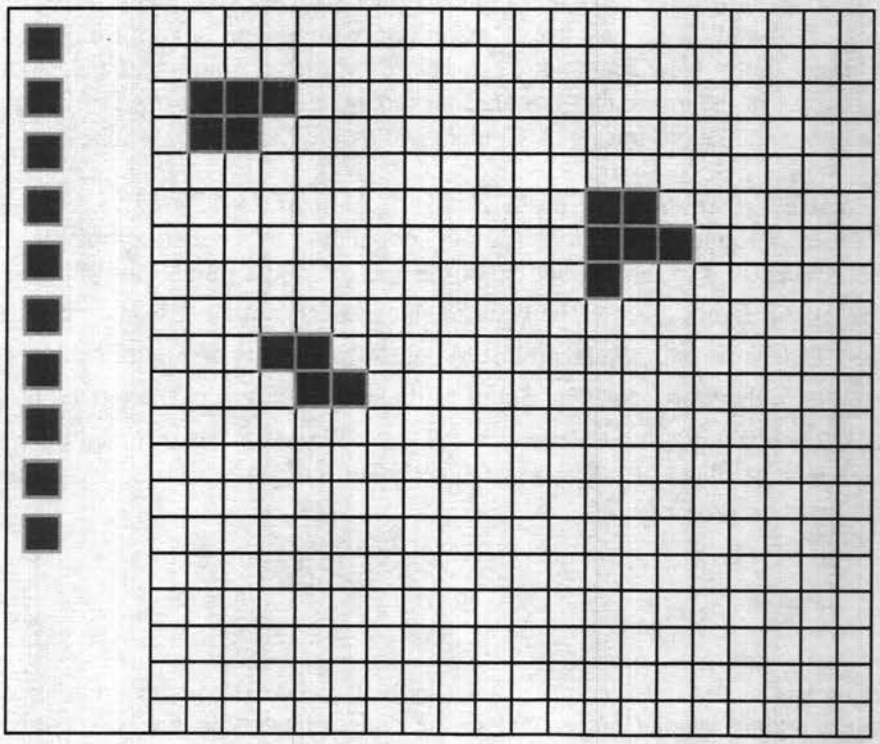


Figure 1. Schematic of the free classification display.

The participants were told that each of the blue squares on the left side of the screen represented a different talker and that the talkers came from different parts of the United States. They were asked to group the talkers based on where they thought the talkers were from. The listeners were allowed to make as many groups as they wanted with as many talkers in each group as they wanted. They did not have to put the same number of talkers in each group and they could listen and move the talkers around as many times as they wanted. No time limit was imposed and the participants were simply instructed to tell the experimenter when they were finished.

Results

A summary of the participants' free classification behavior is shown in Table 1. On average, the participants made 10 groups of talkers, with a range of 3 to 30 and a median of 7. The fact that the median is lower than the mean suggests a skewed distribution, with more listeners making fewer groups and fewer listeners making more groups. The mean number of talkers per group was 9.36, with a range of 1 to 34 and a median of 4. We again have evidence of a skewed distribution with many small groups of talkers being created and fewer large groups of talkers.

Table 1. Summary of the participants' free classification behavior in Experiment 1.

	Mean	Minimum	Maximum	Median
Number of Groups	10	3	30	7
Number of Talkers per Group	9.36	1	34	4

Perceptual Dialect Similarity. The perceptual similarity structure of the six dialect regions was extracted from the free classification data using an additive clustering analysis which produced graphical models of perceptual similarity in tree form (Corter, 1982; Sattath & Tversky, 1977). First, a 6 x 6 dialect similarity matrix was constructed from the free classification data. The matrix reflected the similarity of the dialects as measured by pairwise comparisons of the talkers within each group created by each listener. The value of each cell in the diagonal of the matrix was equal to the number of times that two talkers from the same dialect were put in the same group, summed across all of the listeners. The value of each cell in the off-diagonals of the matrix was equal to number of times one talker from one dialect and one talker from another dialect were put in the same group, summed across all listeners.

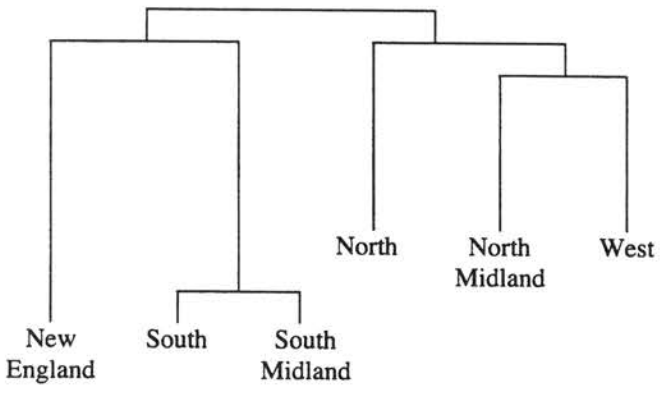


Figure 2. Clustering solution for the listeners in Experiment 1.

To obtain a graphical representation of the perceptual similarity of the six dialects, the 6 x 6 dialect similarity matrix was submitted to ADDTREE (Corter, 1982), an additive clustering analysis.

The clustering solution produced by ADDTREE is shown in Figure 2. In this representation, perceptual similarity is related to the sum of the least number of vertical branches connecting any two dialect nodes and horizontal distance is irrelevant. The clustering analysis revealed three main perceptual clusters: New England, South, and Midwest/West.

Perceptual Talker Similarity. A multidimensional scaling analysis was then carried out to explore the perceptual similarity of the talkers in Experiment 1. A 66 x 66 talker matrix was constructed from the free classification data by assigning to each cell the total number of times a given pair of talkers were put in the same group across all of the listeners. Thus, the 66 x 66 talker matrix reflected the pairwise similarity of all of the talkers. The resulting talker matrix was submitted to a multidimensional scaling analysis and the two-dimensional solution was selected for interpretation and discussion. The dimensions in the two-dimensional space were highly interpretable and the reduction in stress between the two- and three-dimensional solutions was relatively small.

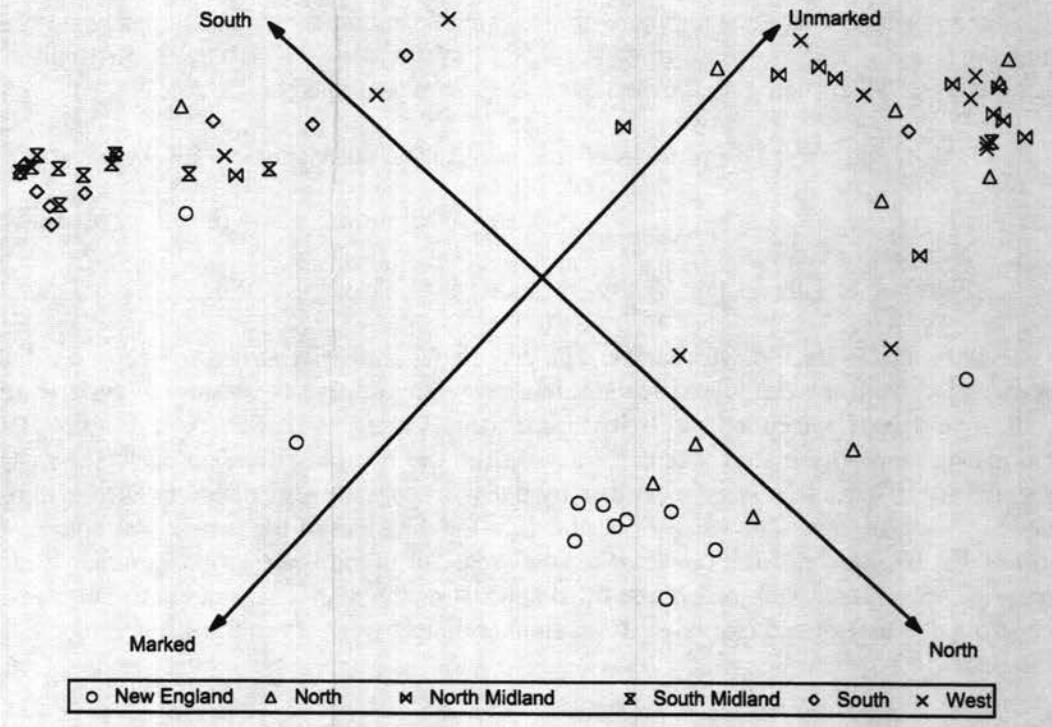


Figure 3. Multidimensional scaling solution for the listeners in Experiment 1.

Figure 3 shows the results of the multidimensional scaling analysis. In this figure, each symbol represents one of the 66 talkers. The three categories that were revealed by the clustering analysis in Figure 2 are also evident in this representation of perceptual similarity. The Southern and South Midland talkers are almost all located in the upper left-hand portion of the space, the New England talkers are almost all in the lower right-hand quadrant, and the rest of the talkers are in the upper right-hand region. Rotation of the space approximately 60° allows for the interpretation of the two perceptual dimensions. From the upper right to the lower left is a dimension related to linguistic markedness, with the marked dialects on the bottom left and the unmarked dialects on the upper right.[13] The second dimension is orthogonal to the first and distinguishes between Southern varieties

[13] The term "linguistic markedness" refers here to the degree to which a given dialect contains phonological variants that are different from the other dialects (Milroy, 2002). The Southern and New England dialects are linguistically marked because they have more features that distinguish them from the other regions. The North

at the upper left and Northern varieties at the lower right.[14] Thus, the two perceptual dimensions that are relevant to the perceptual similarity of the talkers in Experiment 1 are linguistic markedness and geography.

Discussion

The participants in this experiment exhibited a range of free classification strategies, with some listeners making as few as three groups and others as many as 30. The mean number of groups produced was 10, which suggests that listeners can make a relatively large number of distinctions based on regional dialect. When the listeners were asked to provide labels for the groups of talkers they had made after completing the task, they typically provided a list of geographic regions, suggesting that they had followed the instructions and created groups based on their perception of regional varieties of American English. Therefore, while the results of the earlier six-alternative forced-choice categorization tasks seemed to suggest that naïve listeners only have three broad dialect categories, the current results are more consistent with the earlier perceptual dialectology studies (Preston, 1986; Tamasi, 2003) and suggest that listeners try to make finer distinctions between regional varieties when specific labels are not imposed on the task a priori by the experimenter.

Despite the fine-grained classifications made by individual participants, however, the clustering analysis of the aggregate data again revealed three broad dialect categories: New England, South, and Midwest/West. These three perceptual categories correspond directly to the categories revealed by clustering analyses of the confusion matrices in the earlier six-alternative forced-choice categorization tasks using stimuli from the TIMIT corpus (Clopper et al., in press; Clopper & Pisoni, 2004a, b). This result provides converging evidence for the validity of the previous results and confirms that a free classification task can be used to measure the perceptual similarity of regional dialects.

The multidimensional scaling analysis also revealed several novel findings. When the perceptual distances between the 66 talkers in the current study were plotted in a two-dimensional space, the dimensions corresponded to linguistic markedness and geographic location. In particular, the markedness dimension distinguished marked dialect regions, like New England and the South, from unmarked regions, like the North Midland and the West. The geographic dimension, on the other hand, distinguished the Northern varieties, like New England and the North, from the Southern varieties, like the South and South Midland.

The talkers were unevenly distributed in the space defined by the markedness and geographic dimensions. Most of the talkers were located in three locations which represent marked Southern, marked Northern, and unmarked Northern varieties of American English. The unmarked Southern quadrant at the top of Figure 3 is sparsely populated because Southern varieties of American English are typically linguistically marked. In addition, the lower left section of Figure 3 is also relatively empty because marked varieties are typically associated with one end of the geographic continuum, either North or South, with less marked varieties occurring in the middle of the country.

Midland and Western dialects are linguistically unmarked because they have fewer features that distinguish them from the other dialects of American English.

[14] The two dimensions of the space could also be interpreted as a North-South dimension along the original x-axis and an orthogonal East-West dimension along the original y-axis. However, this interpretation would result in a nearly empty South-East quadrant which is somewhat difficult to interpret. In addition, the rotated interpretation is more consistent with the results of the clustering analyses and the results reported below for Experiment 2.

The results obtained in this pilot experiment were promising because they were consistent with previous research on the perception of dialect variation, particularly with respect to the perceptual similarity structure of regional dialects in the United States. In addition, the multidimensional scaling analysis provided new insights into the relevant underlying dimensions of variation for naïve listeners. The next experiment was designed to explore the role of the residential history of the listeners, particularly with respect to geographic mobility and location, in the free classification task. In addition, Experiment 2 used stimulus materials from the new NSP corpus, which more accurately reflect current regional variation in the United States.

Experiment 2

Methods

Listeners. One hundred and six listeners were recruited from the Indiana University community for participation in this experiment. Prior to analyzing the data, 19 participants were removed for the following reasons: seven knew one or more of the talkers by name, one had a parent who was a non-native speaker of English, and 11 did not perform the task as instructed.[15] The remaining 87 listeners were all monolingual native speakers of American English with no reported history of a hearing or speech disorder. The listeners ranged in age from 18 to 25 years old. Both parents of each listener were also native speakers of English. The listeners received $8 for their participation.

As in Chapter 3, the listeners were assigned to four different groups based on their residential history. Twenty-one listeners had lived only in the Midland dialect region and they composed the Non-Mobile Midland group. Twenty-two listeners had lived only in the Northern dialect region prior to attending college at Indiana University in Bloomington and they comprised the Non-Mobile North group. Forty-four listeners had lived in at least two different dialect regions before the age of 18. Twenty-two of these Mobile listeners had parents living in the Midland dialect region at the time of testing and they comprised the Mobile Midland group. The remaining twenty-two Mobile listeners had parents living in the Northern dialect region at the time of testing and they comprised the Mobile North group. Table 2 provides a summary of the residential history of the listeners in Experiment 2.

Table 2. Residential history of the 87 listeners in the free classification task in Experiment 2.

	North	Midland	Total
Mobile	22	22	44
Non-Mobile	22	21	43
Total	44	43	87

Talkers. Forty-eight talkers were selected from the Nationwide Speech Project (NSP) corpus (Chapter 2) for use in Experiment 2. The talkers included four males and four females from each of six dialect regions in the United States: New England, Mid-Atlantic, North, Midland, South, and West. This set of talkers is the same as those used in the forced-choice categorization experiment described in Chapter 3.

[15] Data were excluded for those participants whose responses revealed a greater proportion of between-dialect similarity than proportion of within-dialect similarity. This is equivalent to the proportion of errors being greater than the proportion of correct responses, given all possible responses. It is impossible to determine if these participants were performing the task to the best of their abilities and were simply unable to accurately group the talkers by dialect or if their performance was due to lack of attention to the task. This strict exclusion criteria was used to ensure that the data reflected the best efforts of the participants.

The six dialects included in the NSP corpus were selected because talkers from those regions exhibit systematic differences in their vowel productions. The vowel system of the Northern dialect is characterized by the Northern Cities Chain Shift, which involves a clockwise rotation of the low and low-mid vowels, beginning with the raising and fronting of [æ] (Labov, 1998). The New England dialect also exhibits some properties similar to the Northern Cities Chain Shift (Boberg, 2001). The vowel system of the Southern dialect includes shifts in both the back vowels and the front vowels. The high and mid back vowels are fronted, while the front high and mid tense vowels are centralized and the front high and mid lax vowels are peripheralized (Labov, 1998). Southern speech also exhibits the monophthongization of [ɑy] and [oy] (Thomas, 2001). The vowel system in the "Third Dialect" of American English, which includes New England, the Midland, and the West all show a merger of the low-back vowels [ɔ] and [ɑ] (Labov, 1998). Western speech also exhibits [u] fronting (Labov, Ash, & Boberg, forthcoming; Thomas, 2001). Finally, the Mid-Atlantic dialect is characterized by the raising of [ɔ] and the raising of [æ] in certain lexical contexts (Labov, 1994; Thomas, 2001). The vowel productions of the talkers included in the NSP corpus are discussed in more detail in Chapter 2. It should also be noted that unlike the New England talkers in Experiment 1, who were all r-less, the New England talkers from the NSP corpus used in Experiment 2 were all r-ful.

Stimulus Materials. For each talker, one of the two novel sentences used in the earlier forced-choice categorization task (Chapter 3) was randomly selected for use in the free classification task. The stimulus materials therefore included one novel high probability sentence from the Speech Perception in Noise (SPIN) test (Kalikow, Stevens, & Elliot, 1977) for each of the talkers, for a total of 48 novel sentences. Examples of the high probability sentences are shown in (2). The complete list of stimulus materials used in Experiment 2 is shown in Appendix C.

(2) Ruth poured herself a cup of tea.
The swimmer dove into the pool.
I ate a piece of chocolate fudge.

The stimuli were converted to QuickTime movies using the same methods as in Experiment 1, except that the sound files were leveled to 67 dB using Level16 (Tice & Carrell, 1998) prior to making the stimulus movies.

Procedure. The procedure was identical to that used in Experiment 1, except that only 48 movies were presented to the listeners. Given that both male and female talkers were included in Experiment 2, an additional instruction was also given to the listeners. They were told that they could put males and females in the same group if they thought that they were from the same part of the country.

Results

A summary of the grouping performance in the free classification task for each of the four listener groups is shown in Tables 3 and 4. Table 3 displays descriptive statistics on the number of talker groups created by each of the listener groups. Table 4 shows descriptive statistics on the number of talkers per group for each of the listener groups.

Table 3. Descriptive statistics on the number of talker groups produced by each listener group in Experiment 2.

	Number of Talker Groups			
	Mean	**Minimum**	**Maximum**	**Median**
Mobile North	9.73	3	23	8
Mobile Midland	9.23	4	19	9
Non-Mobile North	7.41	3	14	7
Non-Mobile Midland	7.52	3	15	7
Overall	8.48	3	23	8

Overall, the listeners produced an average of 8.48 groups, with a range of 3 to 23 and a median of 8. The fact that the mean and median are very close suggests that the number of talker groups were symmetrically distributed around the mean. An inspection of the individual data confirms an approximately normal distribution of number of groups produced by the different listeners. A one-way ANOVA on the number of talker groups created by each listener group (Mobile North, Mobile Midland, Non-Mobile North, Non-Mobile Midland) was not significant. An inspection of Table 3 suggests, however, that mobility may be an important variable with respect to grouping behavior. A t-test confirmed that the Mobile listeners created significantly more groups (M = 9.5) than the Non-Mobile listeners (M = 7.5; $t(85) = 2.24$, $p < .05$). T-tests comparing the Mobile and Non-Mobile Northerners and the Mobile and Non-Mobile Midland listeners were not significant. In addition, location was not significant by a t-test.

Table 4. Descriptive statistics on the number of talkers per group produced by each listener group in Experiment 2.

	Number of Talkers Per Group			
	Mean	**Minimum**	**Maximum**	**Median**
Mobile North	6.79	1	36	3
Mobile Midland	6.05	1	31	4
Non-Mobile North	7.64	1	38	4
Non-Mobile Midland	7.86	1	38	5
Overall	7.08	1	38	4

With respect to the number of talkers per group, the listeners created groups containing 7.08 talkers on average, with a range of 1 to 38 and a median of 4. The difference between the mean and the median suggests a skewed distribution with more groups with fewer talkers and fewer groups with more talkers. A one-way ANOVA on number of talkers per group for each listener group (Mobile North, Mobile Midland, Non-Mobile North, Non-Mobile Midland) was not significant. T-tests comparing the Mobile and Non-Mobile listeners and the Midland and Northern listeners were also not significant, suggesting that geographic mobility and location were not significant variables in determining the number of talkers per group produced by the listeners. Thus, while the geographic mobility of the listeners had an affect on their free classification strategy, location did not. In addition, the effect of mobility was only revealed by the comparison of the number of groups of talkers produced by the listeners in each group.

The listeners' performance was also assessed in terms of their ability to accurately group the talkers by dialect in the free classification task. First, for each listener, a "percent correct" score was calculated as the number of times talkers from the same dialect were grouped together out of the total number of possible same-dialect pairings. Second, a "percent errors" score was calculated for each listener reflecting the number of times talkers from different dialects were grouped together out of the

total number of possible different-dialect pairings. Table 5 shows the mean percent correct and error scores for each of the four listener groups.

Table 5. Mean percent correct and mean percent error scores for each of the four listener groups. Standard deviations are shown in parentheses.

	Percent Correct	Percent Errors
Mobile North	26 (17)	19 (16)
Mobile Midland	27 (14)	15 (10)
Non-Mobile North	32 (16)	22 (13)
Non-Mobile Midland	29 (18)	19 (15)
Overall	28 (16)	19 (13)

One-way ANOVAs on the percent correct and percent error scores were computed with listener group (Mobile North, Mobile Midland, Non-Mobile North, Non-Mobile Midland) as the factor. The results of both ANOVAs were not significant, suggesting no differences between the groups in terms of their overall accuracy.

Perceptual Dialect Similarity. Perceptual dialect similarity was assessed using the clustering techniques described in Experiment 1. In order to compare the perceptual similarity spaces of the four listener groups, a 6 x 6 dialect similarity matrix was constructed for each listener group. The four matrices were then submitted to ADDTREE, the additive clustering scheme (Corter, 1982). The resulting graphical models of similarity are shown in Figure 4. As in Figure 2 above, perceptual dissimilarity is a function of the lengths of the least number of vertical bars connecting any two dialect nodes and horizontal distance is irrelevant.

The overall structure of the similarity spaces of the dialects is fairly consistent across the four listener groups. In particular, the Mid-Atlantic and the South are the most distinctive dialects in all four solutions. In addition, the Midland and Western dialects are perceived as highly similar by all four listener groups.

However, some differences in the similarity spaces between the listener groups are clearly visible in the clustering solutions in Figure 4. The perceptual similarity spaces for the Mobile listeners more accurately reflect the phonological characteristics of the dialects, with a tight clustering of the New England, Midland, and Western dialects and greater distances between the Northern, Southern, and Mid-Atlantic varieties. The primary difference in structure between the Mobile North and Mobile Midland listeners is that the Mobile Northern listeners perceived the Northern talkers as more similar to Labov's (1998) "Third Dialect" talkers than the Mobile Midland listeners did.

The perceptual dialect similarity structures for the Non-Mobile listeners are less closely related to the phonological properties of the different dialects. For the Non-Mobile Northern listeners, the most notable feature in the clustering solution is the close relationship between the Northern talkers and the New England, Midland, and Western talkers. This relationship is similar to that reported in Chapter 3 for the six-alternative forced-choice task and may reflect the Non-Mobile Northern listeners' inability to perceive the Northern Cities Chain Shift. The Non-Mobile Midland listeners produced the most unexpected similarity structure, particularly with respect to the high degree of perceptual similarity between the Southern talkers and the Midland and Western talkers.

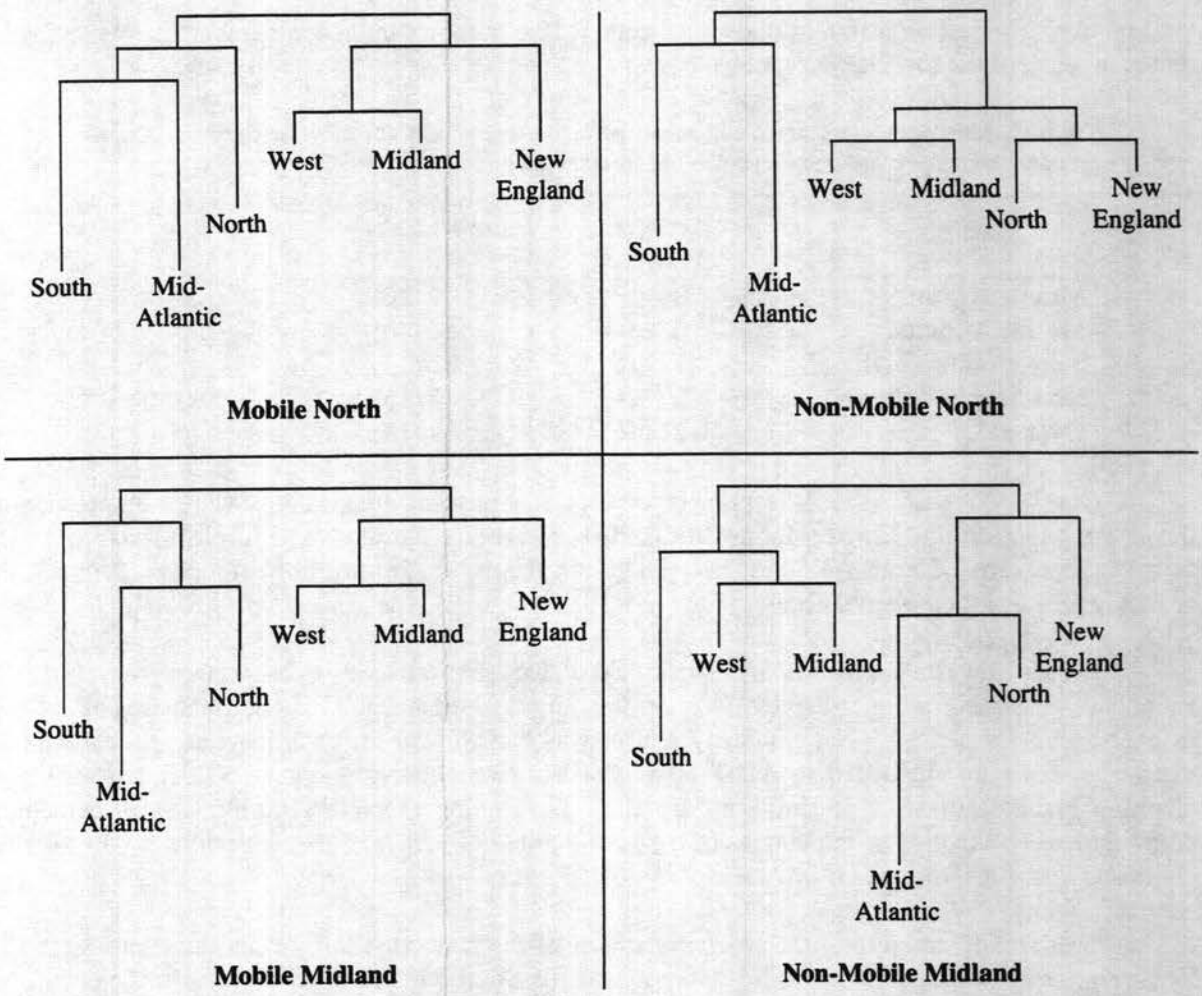


Figure 4. Clustering solutions for the Mobile North, Mobile Midland, Non-Mobile North, and Non-Mobile Midland listener groups in Experiment 2.

One additional pattern that emerges is that the Midland listeners perceived a greater similarity between the Mid-Atlantic and Northern talkers than the Northern listeners. This pattern also reflects the finding that the Midland listeners perceived a greater similarity between the Northern talkers and other marked Northern varieties, such as the Mid-Atlantic dialect than the Northern listeners, who perceived a greater similarity between the Northern and Midland talkers.

Perceptual Talker Similarity. To assess the perceptual similarity of the 48 talkers in Experiment 2, a multidimensional scaling analysis was conducted. As in Experiment 1, a talker similarity matrix was constructed by summing over all of the listeners so that each cell of the matrix indicated the total number of times that that pair of talkers was grouped together in the free classification task. The 48 x 48 talker matrix was then submitted to a multidimensional scaling analysis.

A three-dimensional solution was selected for interpretation and discussion, based on the relatively high interpretability of the dimensions and the fact that stress was much greater for the two-dimensional solution and not greatly reduced for the four-dimensional solution. Figures 5 and 6 show the similarity space produced by the multidimensional scaling analysis in three dimensions. Dimensions 1 and 2 are plotted against each other in Figure 5 and Dimensions 1 and 3 are plotted

against each other in Figure 6. The filled symbols represent male talkers and the empty symbols represent female talkers. For the West, plain X's indicate females and boxed X's indicate males.

Unlike the multidimensional scaling analysis obtained in Experiment 1, the three dimensions produced by the analysis in Experiment 2 are interpretable without rotation. Dimension 1 appears to correspond to markedness, with the linguistically marked dialects on the right and the linguistically unmarked dialects on the left. Most of the Mid-Atlantic and Southern talkers fall on the positive side of Dimension 1. Most of the Midland, New England, and Western talkers, however, fall on the negative, or unmarked, side.

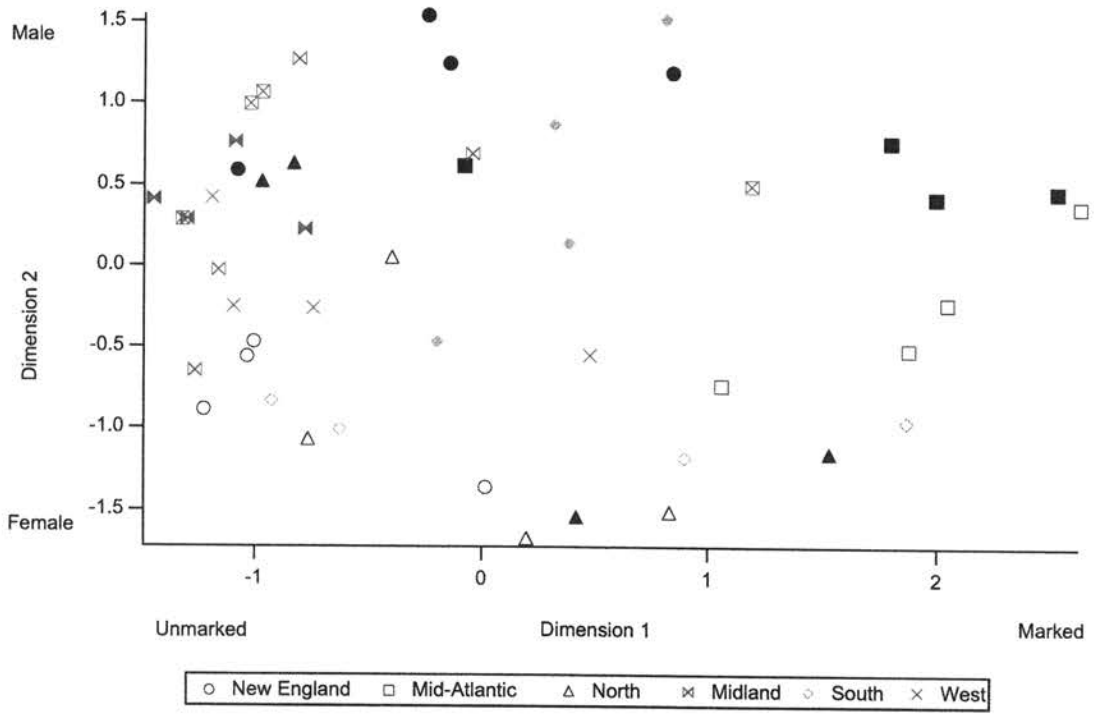


Figure 5. Multidimensional scaling solution for the listeners in Experiment 2 (Dimensions 1 and 2). The filled symbols represent male talkers and the empty symbols represent female talkers. For the West, plain X's indicate females and boxed X's indicate males.

Dimension 2 is related to gender, with nearly all of the male talkers above zero and most of the female talkers below zero. Finally, Dimension 3 can be interpreted geographically, with the South at the top and the North at the bottom. All of the Southern talkers have positive values on Dimension 3, whereas all of the Mid-Atlantic and most of the New England and Northern talkers have negative values.

The space described above and shown in Figures 5 and 6 is the result of combining the responses of all of the listeners in all four of the listener groups. In order to assess differences in perceptual similarity due to residential history, one talker similarity matrix was constructed for each listener group, for a total of four 48 x 48 talker matrices. These four matrices were then submitted to an Individual Differences Scaling (INDSCAL) multidimensional scaling analysis (Carroll & Chang, 1970). The INDSCAL model returns a single group space and subject weights for each matrix. The group space in this case models the perceptual similarity of the 48 talkers and was fixed to be identical to the space produced by the multidimensional scaling analysis based on all of the listeners' data and

plotted in Figures 5 and 6 above.[16] The subject weights indicate the relevance of each of the three dimensions for each listener group. Table 6 shows the normalized weights[17] for each listener group for each dimension.

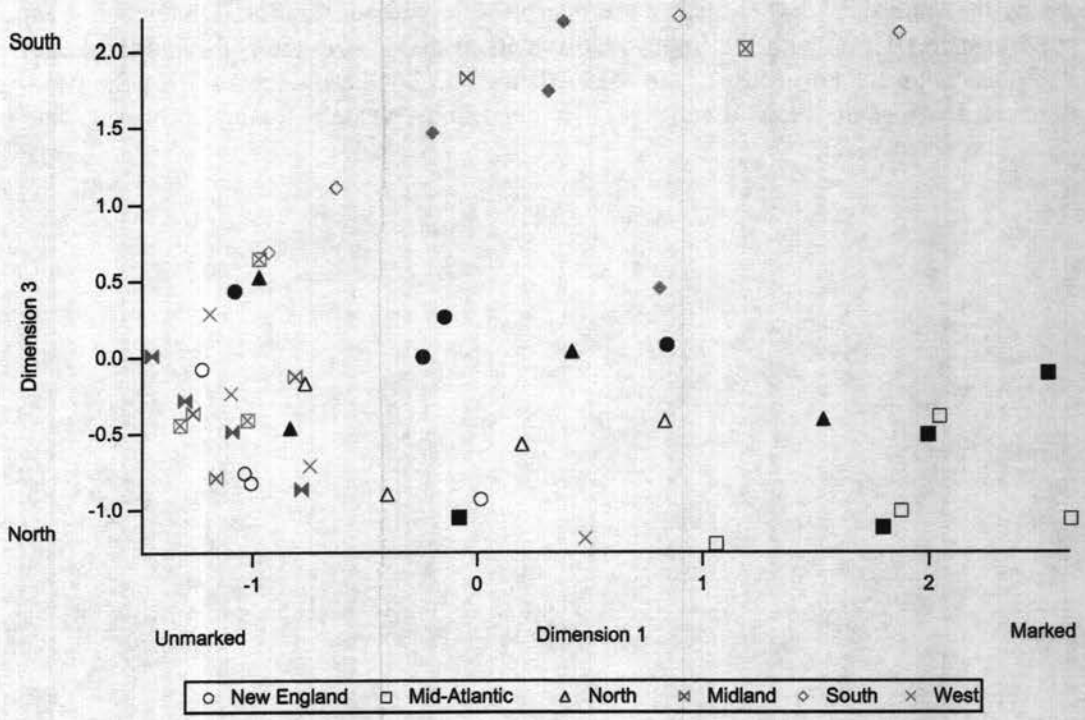


Figure 6. Multidimensional scaling solution for the listeners in Experiment 2 (Dimensions 1 and 3). The filled symbols represent male talkers and the empty symbols represent female talkers. For the West, plain X's indicate females and boxed X's indicate males.

Table 6. Normalized subject weights for each of the four listener groups for each of the three dimensions in the INDSCAL analysis.

	Dimension 1 Markedness	**Dimension 2 Gender**	**Dimension 3 Geography**
Mobile North	0.4076	0.3093	0.2832
Mobile Midland	0.3804	0.3429	0.2767
Non-Mobile North	0.4138	0.2789	0.3073
Non-Mobile Midland	0.4029	0.3184	0.2788
Overall	0.4012	0.3123	0.2865

Overall, the markedness dimension (Dimension 1) received the highest weights, followed by gender (Dimension 2) and geography (Dimension 3). Across all four listener groups, markedness was

[16] An alternative approach to this analysis would be to allow the INDSCAL model to determine both the group space and the subject weights, based on the four listener group matrices. However, that analysis resulted in a noisier similarity space with similar subject weights. Thus, the analysis with the pre-defined similarity space was preferred because the perceptual dimensions were clearer and the subject weights were not substantially different.

[17] The subject weights produced by the INDSCAL analysis are proportional to the overall model fit for each subject. To compare the weights across the listener groups, overall model fit has been factored out by normalizing the subject weights for each group so that they sum to 1.

the most relevant factor in assessing the similarity of the talkers in the free classification task, followed by gender and then geography. The Northern listener groups showed slightly more attention than average to the markedness dimension, while the Midland listener groups showed slightly more attention than average to the gender dimension. The Non-Mobile Northern listeners were most attentive to the geographic dimension, relative to the other listener groups. The listener group differences were quite small, however, suggesting a similar set of relevant perceptual dimensions and attentional weightings across all four groups.

Discussion

Overall performance measured in terms of categorization accuracy was slightly higher in the free classification task (28%) than in the forced-choice task (26%), but the error scores calculated from the free classification data revealed a large percentage of errors (19%) as well. Categorization performance measured in terms of the accuracy and error scores did not differ across the four listener groups. These results are consistent with the results of the six-alternative forced-choice categorization task discussed in Chapter 3, which also did not reveal any main effects of residential history on accuracy. However, in the free classification task, residential history did affect the classification strategies of the listeners. In particular, the Mobile listeners created more talker groups on average than the Non-Mobile listeners. This suggests that personal experience with different regional varieties led to finer-grained perceptual categories for the Mobile listeners.

The Non-Mobile listeners exhibited a tendency for better accuracy scores than the Mobile listeners. Although these differences were not significant, this trend suggests a relationship between accuracy and the number of groups created by the listeners in the free classification task. Across all of the listeners, a significant negative correlation was found between these two measures ($r = -.675$, $p < .001$). To reduce the effects of this dependency between accuracy and classification strategy, the free classification task could be modified so that each listener is required to make the same number of groups (e.g., six or eight). The effects of residential history could then be examined in this more constrained version of the free classification task to determine which effects are due to grouping strategy differences and which are due to differences in overall accuracy.

The perceptual similarity of the six regional dialects was revealed by the clustering analysis. While the overall similarity spaces were consistent across all four listener groups, with Northeast, Southern, and Midwest/West categories, both geographic mobility and location of the listeners affected their performance in the free classification task. In particular, the perceptual similarity structure of the six dialects for the Mobile listener groups were similar to what would be predicted based on the phonological properties of the dialects; Mid-Atlantic and Southern are the most linguistically marked dialects and they were also perceptually the most distinct. The Midland, Western, and New England dialects are the least marked and they clustered tightly together. The Northern dialect is also fairly marked phonologically, but the perception of the Northern talkers seems to be affected by the listeners' location. The Mobile Northern listeners tended to perceive the Northern talkers as more similar to the Midland, Western, and New England talkers, whereas the Mobile Midland listeners heard a greater similarity between the Northern and Mid-Atlantic talkers.

The bias for Northern listeners to hear the Northern talkers as less marked is also evident in the perceptual similarity structure of the six dialects for the Non-Mobile Northern listeners. These listeners perceived the Northern talkers as highly similar to the New England talkers, as well as to the Midland and Western talkers. The poorer perception of the distinctive features of the Northern Cities Chain Shift by Northern listeners was also found in the clustering analysis based on the error patterns in the six-alternative forced-choice task, particularly for the Non-Mobile Northern listeners (Chapter

3). Similar results have also been reported in explicit studies examining the perceptual identification of shifted and unshifted vowels by Northern listeners (Niedzielski, 1999; Rakerd & Plichta, 2003).

Finally, the Non-Mobile Midland listeners showed an unexpected pattern of perceptual similarity with a high degree of similarity between the Southern and Midland talkers. This result is unusual in comparison to the other listener groups, who all perceived the Southern talkers as fairly distinct from the other dialects. This finding is also unexpected, based on the previous six-alternative forced-choice experiments (Chapter 3; Clopper & Pisoni, 2004a) which both revealed high levels of categorization accuracy for Southern talkers and perceptual distinctiveness of Southern talkers, based on clustering analyses, for Non-Mobile Midland listeners.

The high perceptual similarity between Midland and Southern talkers for the Non-Mobile Midland listeners is a direct parallel to the situation for Northern and Midland talkers and Non-Mobile Northern listeners. It is also not surprising in light of some other research on the sociolinguistic situation in the Midland dialect region. Recently, Preston (2002) argued that residents of Southern Indiana have relatively low linguistic security relative to Michigan residents, as assessed by correctness and pleasantness ratings of regional variation.[18] Participants from Michigan rate their own speech as both highly pleasant and highly correct. Participants from Southern Indiana, however, rate their own speech as highly pleasant, but no more or less correct than most of the United States. In addition, they make a strong distinction between the correctness of their own speech and that of their southern neighbor, Kentucky. Thus, when asked explicitly to categorize talkers as coming from the Midland or the South in a forced-choice categorization task, Midland listeners may pay more attention to the differences between their own speech and that of the "incorrect" Southerners in making their responses. However, when they are simply asked to group the talkers by region without any labels or specific categories in the free classification task, those same Midland listeners may perceive their Southern neighbors as being more similar to themselves.

Thus, with respect to judgments of perceptual dialect similarity, both geographic mobility and location appear to be relevant in shaping naïve listeners' perceptions. Geographic location is important for distinguishing between local dialects, particularly with respect to the Northern listeners' misperception of the Northern Cities Chain Shift. Mobility increases listeners' familiarity with other dialects, which leads to more perceptual categories and better discrimination of local dialects. Specifically, the Mobile Northern listeners perceived a greater distinction between the North and the Midland than the Non-Mobile Northerners did. Similarly, the Mobile Midland listeners perceived a greater difference between the Midland and the South than the Non-Mobile Midland listeners did.

The perceptual similarity structures revealed by the free classification task differ in several important ways from those produced from the six-alternative forced-choice confusion data. In Chapter 3, New England was clustered most closely with the Mid-Atlantic in the clustering solutions for all of the listener groups. However, the perceptual similarity of the New England and Mid-Atlantic talkers was asymmetric; Mid-Atlantic talkers were categorized as New England more often than New England talkers were categorized as Mid-Atlantic. Based on these findings, we suggested that the listeners had only one category for Northeastern speech and that this category reflected the phonological variables associated with the Mid-Atlantic talkers. This interpretation is supported by the free classification data which show a high degree of perceptual similarity between New England and the other "Third Dialect" regions, including the Midland and the West. Participants were also slightly

[18] Linguistic security refers to participants' beliefs about the prestige of their own dialect, relative to other language varieties. Linguistically secure participants believe that their own dialect is prestigious, whereas linguistically insecure participants believe that their own dialect is less prestigious than some other standard variety.

more accurate overall in the free classification task than in the forced-choice task. The labels provided in the six-alternative task led to a response bias that was eliminated in the free classification task, allowing more accurate perceptual clusters to emerge.

A second major difference between the perceptual similarity spaces produced from the two different tasks is the perceived similarity between the Midland and Southern talkers for the Midland listeners. In Chapter 3, we found that the Southern talkers were categorized more accurately than most of the other talker groups by all four listener groups. In addition, the Southern dialect was revealed to be distinctive from the other dialects in the clustering analyses based on the stimulus-response confusions for all of the listener groups. In the free classification experiment, however, the Non-Mobile Midland listeners behaved like the Non-Mobile Northern listeners in that they failed to distinguish talkers from their own region from talkers from a neighboring region. Specifically, while the Non-Mobile Northern listeners perceived a greater similarity between the Northern and Midland talkers than the other listener groups, the Non-Mobile Midland listeners perceived a greater similarity between the Midland and Southern talkers than the other listener groups. This finding suggests that the labels provided in the forced-choice experiment (Chapter 3) forced the Non-Mobile Midland listeners to make distinctions between Midland and Southern talkers that they did not find relevant in the free classification task.

The multidimensional scaling analysis revealed a perceptual similarity space of the talkers with three dimensions: linguistic markedness, gender, and geography. Together, the markedness and geography dimensions define a space in which the three broad dialect categories can be distinguished (see Figure 6). The Southern dialect occupies the Marked-Southern quadrant of the space, the Mid-Atlantic and Northern dialects occupy the Marked-Northern quadrant of the space, and the Midland, Western, and New England dialects occupy the Unmarked-North quadrant. The Unmarked-South quadrant is fairly empty, because Southern varieties of American English are typically marked.

The fact that gender emerged as a relevant dimension in the analysis is interesting because the listeners were told to ignore it in making their groups. This instruction may be difficult to follow, however, because gender and dialect are known to interact. For example, women typically lead phonological change, which means that women might produce more of some variables than men. On the other hand, women also tend to avoid stigmatized forms, which means that they might produce fewer of some other variables than men (Labov, 1990). Thus, the fact that gender emerges as an important dimension may simply reflect the listeners' sensitivity to the complex interaction between gender and regional dialect.

The finding that naïve listeners attend to the interaction between linguistic (e.g., phonological) and indexical (e.g., gender or dialect) information in the speech signal is also consistent with previous findings which suggest that speech perception is a talker-contingent process. For example, Nygaard, Sommers, and Pisoni (1994) found that speech intelligibility in noise was more accurate when the talkers were familiar to the listeners than when they were unfamiliar, suggesting an integration of phonological and talker-specific information in speech perception. The results of the present study suggest that naïve listeners can integrate several different kinds of talker-specific and linguistic information in order to make judgments about the dialect of a given talker.

The four listener groups consistently attended most to the markedness dimension in assessing the similarity of the talkers in the free classification task. In general, gender was the next most important dimension, followed by geography. However, a few group differences were revealed by the INDSCAL analysis. In particular, the Northern listeners relied more on the markedness dimension than the Midland listeners and the Midland listeners relied more on the gender dimension than the Northern listeners. Finally, the Non-Mobile Northern listeners attended more to the geography

dimension than the other listener groups and also relied less on gender than on geography. This last result is difficult to interpret because Non-Mobile Northern listeners do not make perceptual distinctions between Northern and Midland talkers, which might suggest that they attend less to the geographical dimension instead of more. Further research is needed to explore the role of these three perceptual dimensions in classification behavior.

General Discussion

Across the two free classification experiments, the listeners made 8-10 groups of talkers on average. The slightly smaller number of talker groups in Experiment 2 is probably a reflection of the fewer number of stimuli presented to the listeners (48 in Experiment 2 vs. 66 in Experiment 1). It was surprising that the participants made as many groups as they did, because previous forced-choice categorization studies suggested that listeners could only accurately distinguish three regional varieties of American English (Chapter 3; Clopper et al., in press; Clopper & Pisoni, 2004b). However, the results of the free classification experiments suggest that naïve listeners can make finer distinctions between regional dialects. That is, their knowledge of dialect variation may include more gradient categories than previously reported based on the results of the forced-choice categorization tasks. This finding suggests that the labels provided to the listeners in the forced-choice task may not accurately reflect the categories for regional dialects that naïve listeners actually have and can use in explicit tests of dialect classification. Further research is needed to determine if the labels and response alternatives are inconsistent with naïve listeners' category names, the regions they correspond to, or phonological properties associated with the labels.

Perceptual Dialect Similarity

The clustering analyses of the free classification results revealed three main perceptual dialect clusters in both Experiment 1 and Experiment 2: Northeast, South, and Midwest/West. These clusters are consistent with the perceptual categories revealed by the error patterns in forced-choice categorization tasks (Chapter 3; Clopper et al., in press; Clopper & Pisoni, 2004b) and with phonological descriptions of the major varieties of American English. Labov (1998) described the three main regional dialects of English in North America as Northern, Southern, and the "Third Dialect," which includes New England, the Midland, and the West. These three primary regional dialects also have some historical significance, because they were also described by Krapp (1925) as Eastern, Southern, and Western (or General) American and reflect early migration and settlement patterns in the United States (Carver, 1987).

The differences in the clustering solutions between the two experiments can be attributed to several methodological factors, including the different stimulus materials. First, the specific regional labels applied to the talkers in the TIMIT corpus differed from those used to describe the NSP talkers. The most important difference is that Mid-Atlantic talkers were not included in Experiment 1, so perceptual differences between Mid-Atlantic and New England talkers could not be assessed. Second, even those labels which were identical across the two studies did not correspond to the same geographic area. Thus, the Southern region in the NSP includes both the South and the South Midland TIMIT regions and the TIMIT North Midland region corresponds most closely to the NSP Midland area. Third, only male talkers were used in Experiment 1, whereas both male and female talkers were used in Experiment 2. As discussed above, further research is needed to determine the relationship between gender and dialect variation in the perception of dialect similarity. Finally, the New England talkers in Experiment 1 were all r-less, which has been shown to be a salient cue for naïve listeners (Clopper & Pisoni, 2004b). However, none of the New England talkers in Experiment 2 were r-less, which means that the listeners needed to rely on other phonological properties, such as vowel quality,

to distinguish the New England talkers from the other talkers in this task. This difference in r-lessness between the two groups of talkers is probably the major factor responsible for the relatively greater perceptual distinctiveness of the New England talkers in Experiment 1 than Experiment 2.

One additional methodological difference between the first and second experiments was the residential history of the listeners. In Experiment 1, the listeners were mixed with respect to residential history, whereas in Experiment 2, the listeners were grouped based on their geographic mobility and location. Given that we found significant effects of residential history in Experiment 2, the differences between the perceptual similarity spaces in the two experiments might be a result of collapsing across residential history in Experiment 1. It is more likely, however, that the differences are due to the stimulus materials used in each experiment, given the overall consensus between the listener groups in Experiment 2.

As discussed above, the clustering solution revealed by the free classification data in Experiment 1 is consistent with the results of the clustering analyses performed in the earlier forced-choice tasks using the TIMIT talkers (Clopper et al., in press; Clopper & Pisoni, 2004b). However, a comparison of the clustering solutions in Experiment 2 to those in the earlier forced-choice task using the NSP talkers (Chapter 3) reveals a difference in perceptual similarity due to the task demands. Specifically, New England and Mid-Atlantic talkers were perceptually more similar in the forced-choice task than in the free classification task. This difference is presumably due to an effect of the closed set of verbal labels in the forced-choice task which influenced the listeners' responses by imposing an external category structure on the talkers.

The responses of the listeners in the forced-choice task revealed a strong stimulus-response asymmetry between the Mid-Atlantic and New England dialect categories (Chapter 3). The listeners were approximately twice as likely to miscategorize a Mid-Atlantic talker as a New Englander than vice versa. We concluded from these results that the general unfamiliarity of the listeners with the New England dialect led to a mismatch between the category label (i.e., "New England") and a representation of the actual phonological properties of New England speech. In the free classification task, the effects of this mismatch were minimized because the listeners did not need to apply verbal labels to the talkers in order to perform the task, but could merely attend to the phonological properties of the different talkers.

Perceptual Talker Similarity

The perceptual similarity of the talkers was explored in two free classification experiments using multidimensional scaling techniques. In Experiment 1, two dimensions of perceptual similarity were revealed: linguistic markedness and geography. In Experiment 2, the same two dimensions were found, along with a third dimension related to gender. Gender was not an important perceptual dimension in Experiment 1 because all of the talkers were male. Thus, the dimensions of the perceptual structure revealed by the free classification task was replicated with the two different sets of stimulus materials.

The finding that markedness emerged as the most important dimension in the free classification task is not surprising. Linguistic markedness also emerged in the clustering analyses, with marked dialects appearing as the most distinctive dialects (New England, South Midland, and South in Experiment 1 and Mid-Atlantic and South in Experiment 2). In the earlier six-alternative tasks, performance measured in terms of accuracy was typically greater for linguistically marked dialects than unmarked dialects. Clopper and Pisoni (2004b) reported higher overall performance on New England and Southern talkers and lower overall performance on Western and North Midland talkers. In Chapter 3, we found higher performance on the Southern and Mid-Atlantic talkers than on

the Western and New England talkers. Performance was higher than expected on the unmarked Midland talkers in Chapter 3 due in part to a strong response bias in favor of the Midland. Finally, even the perceptual dialectology studies showed some effects of markedness; linguistically marked regions such as the South and the Northeast were among the most commonly indicated dialect regions in the map-drawing and free classification tasks (Preston, 1986; Tamasi, 2003).

The geographic dimension is also consistent with previous research on the perception of variation by naïve listeners. In particular, Preston and his colleagues (Plichta & Preston, 2003; Preston, 1993) have found that naïve listeners can distinguish between talkers using a North-South continuum. Naïve listeners were able to make broad North vs. South distinctions in a dialect identification task using speech samples obtained from nine male talkers from nine cities between Saginaw, Michigan and Dothan, Alabama (Preston, 1993). Listeners made even finer distinctions between utterances synthesized to vary in degree of [ɑy] diphthongization using the same nine-step geographic range (Plichta & Preston, 2003). The finding that geography emerged as an important dimension in the perception of the similarity of regional varieties of American English is consistent with the results of these earlier studies.

As discussed above, gender and dialect interact in speech production, so the appearance of gender as the third relevant dimension in the multidimensional scaling analysis in Experiment 2 may be rooted in production differences. Clopper et al. (in press) reported some differences due to gender in perceptual similarity of dialects as assessed by clustering analyses of error patterns in the six-alternative forced-choice task. In particular, the female Northern talkers were perceived as being more similar to the New England talkers, whereas the male Northern talkers were perceived as being more similar to the North Midland and Western talkers. These results suggest a potential relationship between gender, production of specific phonological properties such as the Northern Cities Chain Shift, and the perceptual similarity of regional dialects. Further research on the role of talker gender in the perception of dialect variation is needed to fully understand how the interaction of these variables in speech production is related to their interaction in perception.

Linguistic Experience and Free Classification

Residential history was found to be an important factor in the performance of the listeners in Experiment 2. While it did not affect overall accuracy, it did affect general grouping strategy in the free classification task and the perceptual similarity of the dialects and the talkers. In particular, mobility was found to be relevant to classification strategy, with the Mobile listeners making more talker groups than the Non-Mobile listeners. Mobility also affected perceptual similarity of the dialects, as revealed by the clustering analysis. The Mobile listeners perceived greater distinctions between geographically local dialects than the Non-Mobile listeners. Specifically, the Mobile Northern listeners perceived a greater difference between Midland and Northern talkers than the Non-Mobile Northern listeners did. Similarly, the Mobile Midland listeners perceived a greater difference between Midland and Southern talkers than the Non-Mobile Midland listeners did. In terms of the acquisition of reliable categories, geographic mobility provides naïve listeners with the opportunity to interact with talkers from different parts of the country and to accurately assign dialect labels to those individuals. Non-mobile listeners, however, have fewer opportunities to make connections between talkers and their residential history and, by extension, their appropriate dialect category label. This difference between exposure to talkers and their dialect labels led to more accurate perceptual similarity between dialect categories for the mobile listeners than the non-mobile listeners (see Barsalou, 1985).

Geographic location was also a factor in the perceptual similarity of the different dialects. The Northern listeners perceived a greater overall similarity between the Northern talkers and the Midland talkers than the Midland listeners did. Similarly, the Midland listeners perceived the Midland and Southern talkers as more similar than the Northern listeners did. In addition, location appeared to be an important variable with respect to the attention that the listeners paid to the different dimensions in assessing the similarity of the talkers in the free classification task. While the overall emphasis on the perceptual dimensions was similar across the four listener groups, the Northern listeners relied more on the markedness dimension than the Midland listeners did and the Midland listeners relied more on the gender dimension than the Northern listeners did.

Conclusions

The free classification task revealed several new insights about the underlying perceptual similarity structure of regional varieties of American English for naïve listeners. First, the listeners in both experiments made more groups of talkers than predicted, suggesting that they are able to make fine-grained distinctions between regional dialects. Second, the clustering analyses revealed parallel patterns of similarity structure to those revealed by the earlier forced-choice tasks (Chapter 3; Clopper et al., in press; Clopper & Pisoni, 2004b). These findings provide converging evidence for the primary perceptual dialects of American English: Northeast, South, and Midwest/West. In Experiment 2, we found that New England talkers were perceived as being more similar to Midland and Western talkers, with whom they share several phonological properties, than in the previous forced-choice task (Chapter 3), in which they were perceived as being most similar to their geographic neighbors from the Mid-Atlantic region. This result suggests that the labels provided by the experimenter in the forced-choice task may lead to some response biases that can be reduced by using a free classification task. Finally, the multidimensional scaling analyses revealed three primary dimensions of perceptual organization of dialect variation: linguistic markedness, geography, and gender. All three of these dimensions have both linguistic and social significance, suggesting that naïve listeners build categories for regional variation based on appropriate sociolinguistic dimensions. The free classification task used in the present study is therefore a promising new methodology in the study of the perception of dialect variation by naïve listeners.

References

Ayers, E.L., Limerick, P.N., Nissenbaum, S., & Onuf, P.S. (1996). *All Over the Map: Rethinking American Regions.* Baltimore: Johns Hopkins University Press.

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Boberg, C. (2001). The phonological status of Western New England. *American Speech, 76*, 3-29.

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 238-319.

Carver, C.M. (1987). *American Regional Dialects: A Word Geography.* Ann Arbor, MI: University of Michigan Press.

Clopper, C.G., Conrey, B.L., & Pisoni, D.B. (in press). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology.*

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.

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

Evans, B.G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115*, 352-361.

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

Imai, S. (1966). Classification of sets of stimuli with different stimulus characteristics and numerical properties. *Perception and Psychophysics, 1*, 48-54.

Imai, S., & Garner, W.R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology, 69*, 596-608.

Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Krapp, G.P. (1925). *The English Language in America*. New York: Frederick Ungar.

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. (1994). *Principles of Linguistic Change: Internal Factors*. Malden, MA: Blackwell.

Labov, W. (1998). The three dialects of English. In M.D. Linn (Ed.), *Handbook of Dialects and Language Variation* (pp. 39-81). San Diego: Academic Press.

Labov, W., Ash, S., & Boberg, C. (forthcoming). *Atlas of North American English*. New York: Mouton de Gruyter.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in Sound: The Cognitive Psychology of Human Audition* (pp. 146-198). Oxford: Clarendon.

Medin, D.L., Wattenmaker, W.D., & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242-279.

Milroy, L. (2002). Social networks. In J.K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 549-572). Malden, MA: Blackwell.

Murphy, G.L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

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

Nosofsky, R.M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 54-65.

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

Plichta, B., & Preston, D.R. (2003). The /ay/s have it: Stereotype, perception, and region. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

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. (1999). *Handbook of Perceptual Dialectology*. Amsterdam: 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.

Rakerd, B., & Plichta, B. (2003). More on perceptions of /ɑ/ fronting. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*, 319-345.

Tamasi, S.L. (2003). *Cognitive Patterns of Linguistic Perceptions*. Doctoral dissertation, University of Georgia, Athens.

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

Thomas, E.R. (2002). Sociophonetic applications of speech perception experiments. *American Speech, 77*, 115-147.

Tice, R., & Carrell, T. (1998). Level16 (Version 2.0.3) [Computer Software]. Lincoln, NE: University of Nebraska.

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology, 18*, 31-48.

Van Bezooijen, R., & Ytsma, J. (1999). Accents of Dutch: Personality impression, divergence, and identifiability. *Belgian Journal of Linguistics, 13*, 105-129.

Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D. R. Preston (Ed.), *Handbook of Perceptual Dialectology* (pp. 345-358). Philadelphia: John Benjamins.

CHAPTER 5: THEORETICAL IMPLICATIONS[19]

Introduction

The research described in this dissertation examined the perceptual classification of regional varieties of American English by naïve listeners, using two experimental methodologies developed in the field of cognitive science to measure perceptual similarity: forced-choice categorization and free classification. The listeners' performance on the forced-choice task was poor, but statistically above chance, suggesting that naïve listeners have reliable perceptual categories for regional dialect variation. Detailed analyses of the patterns of errors revealed three main perceptual dialects: Northeast, South, and Midwest/West. These perceptual dialects were found to be influenced by the residential history of the listeners with the Non-Mobile Northern listeners perceiving a greater similarity between the Northern and Midland talkers than the other listener groups. In addition, the Mobile Northern listeners perceived a greater difference between the marked Southern and Mid-Atlantic dialects and the less marked Midland and Western dialects than the Midland listeners.

The results of the free classification task also revealed three primary perceptual dialects: Northeast, South, and Midwest/West. An analysis of the free classification behavior of the listeners revealed three perceptual dimensions related to talker similarity and regional dialect: linguistic markedness, geography, and gender. Residential history of the listeners was also found to be an important variable that affected the judgments of perceptual similarity of the talkers in the free classification task. Non-mobile listeners perceived greater similarity between themselves and their neighbors than the mobile listeners. Taken together, these new findings on dialect perception have implications for the fields of sociolinguistics, speech science, and theoretical phonology.

Implications for Sociolinguistics

Based on the results of an earlier six-alternative forced-choice perceptual dialect categorization task, Clopper and Pisoni (2004b) argued that naïve listeners have three main perceptual dialect clusters: Northeast, South, and Midwest/West. These initial results were replicated with a new corpus in a six-alternative forced-choice task (Chapter 3) and with two different sets of stimulus materials in a free classification task (Chapter 4). As discussed above, these three perceptual dialects are also consistent with descriptions of regional dialects in the United States based on phonological and lexical analyses of variation. In particular, Labov (1998) described the major varieties of American English as North, South, and the "Third Dialect," which includes the Midland and West, based on acoustic analyses of the vowel systems of talkers from all over the United States. Carver (1987) described the dialects of American English in terms of Southern and Northern varieties, based on lexical variation. Finally, as early as 1925, Krapp discussed Eastern, Southern, and Western (or General) varieties of American English, based on impressionistic transcriptions of pronunciation differences. The perceptual research reported above suggests that these major dialects, as defined by sociolinguists, are also salient and stable perceptual dimensions of speech for naïve listeners.

The free classification task also uncovered several dimensions of perceptual similarity related to dialect variation (Chapter 4). These dimensions included linguistic markedness, geography, and gender and suggest that the naïve listeners were attending to both linguistic and non-linguistic aspects of the speech signal when they were asked to classify the talkers by dialect. The fact that markedness emerged as the first perceptual dimension is consistent with previous research which revealed naïve

[19] Sections of this chapter will appear in Clopper, C. G., & Pisoni, D. B. (in press). Perception of dialect variation. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception*. Oxford: Blackwell.

listeners' sensitivity to phonological cues to dialect affiliation, such as r-lessness (Clopper & Pisoni, 2004b) and [ɑy] monophthongization (Plichta & Preston, 2003). The geographical dimension has also been found to be salient in previous research on the perception of dialect variation (Preston, 1993) and reflects perhaps the most important geographic dimension of linguistic variation in the United States. Historically, regional dialects spread across the country from East to West in three broad bands (Carver, 1987), creating the Northern, Midland, and Southern dialects. The present findings demonstrate that naïve listeners are sensitive to this important aspect of regional variation in the United States.

Finally, gender has been found to interact with regional dialect in a number of ways (Labov, 1990). First, women tend to be more advanced then men in terms of the production of phonological change, so that some variables occur more in women's speech than in men's. However, women also tend to avoid stigmatized forms, so that other variables occur more often in men's speech than in women's. Clearly, more research is needed to explore the interaction between regional dialect and gender in speech production and speech perception. However, the current results are consistent with those reported by Clopper, Conrey, & Pisoni (in press) and suggest that naïve listeners have some knowledge of the interaction between gender and dialect in speech production.

The results of the present study also revealed effects of residential history on the perception of dialect variation. While we did not find any overall effects of accuracy due to mobility or location, both variables affected the perceptual similarity of the different regional varieties. The most striking result was the apparent "shift deafness" of the Non-Mobile Northern listeners who did not perceive the phonological differences between the Northern and Midland talkers. Chapter 2 confirmed the presence of the Northern Cities Chain Shift in the speech of the Northern talkers in the NSP corpus, but the Non-Mobile Northern listeners in the forced-choice categorization and free classification tasks did not appear to be sensitive to this vowel shift. Evidence for this "shift deafness" has also been found in a vowel matching task (Niedzielski, 1999) and a word recognition task (Rakerd & Plichta, 2003), using synthetic vowel stimuli with a range of shifted and unshifted vowels. The results of all three studies demonstrate a significant perceptual bias of Non-Mobile Northern listeners to perceive their own speech (and the speech of those around them) as containing unshifted vowels, when in fact the Northern Cities Chain Shift is quite prevalent in the production of talkers from a large geographic area in the Northern United States (Chapter 2; Labov, Ash, & Boberg, forthcoming). Further research on the perception of dialect differences by this group of listeners is needed to determine at which level of encoding and representation the misperceptions are occurring.

The "shift deafness" exhibited by the Northern listeners is perhaps the most striking example of a difference in perception between naïve listeners and trained sociolinguists. The results of the current research as well as previous research by Plichta and colleagues (e.g., Plichta & Preston, 2003; Rakerd & Plichta, 2003) suggest that naïve listeners may not attend to all of the linguistic variants that are present in their own speech and the speech of their neighbors. Phillips (2001) and Pierrehumbert (2001) both provide evidence for lexical frequency effects on phonological sound change. In some cases, high frequency words are the first to shift, followed by lower frequency words. In other cases, high frequency words are resistant to change and shifts occur first in low frequency words. Additional research is needed to determine not only which properties of different dialects are most salient for naïve listeners, but also to explore the effects of lexical frequency and phonological change on the perception of linguistic variation.

Implications for Speech Science

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 (Klatt, 1989; Stevens, 1996). Traditional models of speech perception have assumed that variation is stripped off early in a process of normalization so that the meaningful symbolic content of the signal can be recognized (see Pisoni, 1997). This foundational assumption is central to all of the traditional abstractionist views of speech and language as symbolic systems, in which variation is treated as irrelevant noise (e.g., Chomsky & Halle, 1968; Joos, 1948).

Researchers have only recently begun to moderate the traditional symbolic view of language and to investigate and understand the contributions of linguistic variability in human speech perception. An extensive literature on the role of talker variability and talker-specific information in speech perception over the last decade suggests that indexical properties of speech are perceived and encoded by listeners in laboratory-based psycholinguistic processing 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, although it seems to be more poorly or less robustly encoded than other types of talker-specific information.

The participants in the six-alternative forced-choice task were only able to correctly categorize 26% of the unfamiliar talkers by dialect under good listening conditions. In contrast, naïve adults and children can identify the gender of unfamiliar talkers with high levels of accuracy, even with whispered speech or low-pass filtered stimuli (Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Bennett & Montero-Diaz, 1982). Naïve adult listeners have also demonstrated high levels of accuracy in the identification of familiar talkers in degraded listening conditions (Remez, Fellowes, & Rubin, 1997; Sheffert, Pisoni, Fellowes, & Remez, 2002). Thus, whereas talker dialect is only poorly categorized under good listening conditions, both talker identity and gender are accurately categorized in degraded listening conditions. Taken together, the results of these studies suggest that talker identity and gender may be more robustly represented by naïve listeners than talker dialect.

The dimensions of perceptual similarity revealed by the free classification task suggest that naïve listeners are attending to relevant linguistic and social dimensions and are using these attributes in making their classifications. However, the low levels of accuracy suggest that the representations that naïve listeners have for regional dialect variation are incomplete. One possible explanation for this incompleteness of dialect category representations is that naïve listeners may not always have enough reliable non-linguistic information to associate a given talker or utterance with a specific dialect category. Whereas a listener can frequently use non-linguistic cues to identify a talker's gender and in many social settings unfamiliar talkers readily provide their own name to their listeners, it is far less common for an unfamiliar talker to provide a detailed description of his or her regional background. Presumably, dialect categories must be constructed based only on utterances produced by talkers for which the regional background is known.

This issue of the listeners' familiarity with the different dialects is particularly relevant to the classification of the New England talkers. In the forced-choice task, the listeners made more errors on the New England talkers than on any other talker group. In addition, they displayed a response bias against the "New England" response, except for Mid-Atlantic stimuli which were often (25%) misidentified as New England. In the free classification task, however, the New England talkers were perceptually more similar to the Midland and Western talkers, with whom they share some

phonological attributes. These findings suggest that the listeners in the current study have relatively poor representations of what New England talkers sound like, although they may have a category label for "New England." That is, these listeners believe that New England speech is supposed to be its own category (c.f., Preston, 1986), but they have not had enough exposure to talkers from New England to have formed robust representations of New England phonology.

A listener's familiarity with different linguistic varieties may also affect cross-dialect intelligibility. One early study of cross-dialect speech intelligibility found that within-dialect intelligibility was higher than between-dialect intelligibility (Mason, 1946), which suggests that unfamiliar linguistic variants may interfere with spoken language processing and lexical access. These earlier findings are consistent with the recent proposal that talker-specific information in speech is encoded and represented as part of the normal speech perception process (Pisoni, 1997). Additional research is needed to investigate perceptual learning of new dialect categories by naïve listeners and the effects of dialect variation on speech intelligibility in noise and under other degraded listening conditions. These kinds of studies will provide additional insights into the role of dialect variation in speech perception and spoken language processing.

Implications for Theoretical Phonology

Research on the perception of dialect variation also has some important implications for the field of theoretical linguistics. Like many speech scientists, 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 usually assume a one-to-one mapping between idealized symbolic phonological forms in the mental lexicon and phonetic outputs in production. The research reported in this dissertation, however, suggests that phonological variation is an important property of human speech perception and models of phonological systems should be able to account for these physically and psychologically real aspects of human language performance.

In acquiring the grammar of any language, the language user is exposed to enormous variation across different talkers and social settings. However, naïve listeners quickly learn which aspects of this variation are linguistically significant (i.e., allophonic) and which are due to other sources of variation, such as those described by Klatt (1989; Jusczyk, 1997). Research on dialect perception suggests that listeners also learn which of the non-allophonic variation is socially marked and which is not. That is, language learners acquire a linguistic system in which some kinds of variability are systematic (e.g., dialect variation and coarticulatory effects). A language user's linguistic competence includes knowledge about what kinds of variation are permissible and which are not.

Phonological models must also provide an account for the systematic sources of variation that human listeners perceive, encode, and represent. The results of Niedzielski's (1999) vowel matching experiment suggest that naïve listeners have representations of both their own phonological variants, as exhibited by their production, as well as some baseline variants, as demonstrated by their ability to reliably select unshifted vowel productions in laboratory perception tasks. While it remains an empirical question as to how different variants of a single vowel category are perceived, encoded, and represented, a complete model of a phonological system must account for naïve listeners' knowledge about permissible phonological variation in their native language.

Experience with linguistic variation has been shown to play an important role in dialect categorization performance (Chapter 3; Chapter 4; Clopper & Pisoni, 2004a). Exemplar models of language processing are particularly well suited to account for frequency- and experience-based

phenomena (Bybee, 2001; Johnson, 1997). For example, Goldinger (1996) conducted a series of behavioral studies demonstrating the specificity of memory representations of spoken words. He argued that exemplars of spoken words are accessed in word identification tasks and suggested that an exemplar model of the lexicon can account for both the nature of representations in memory as well as online perception of novel stimuli. Based on findings from several experimental procedures, he proposed an exemplar-based model of the human lexicon in which words are stored as episodic traces and abstract lexical entries are unnecessary (Goldinger, 1998).

More recently, however, the proposal that abstract units can be replaced entirely by exemplar models (often instantiated computationally as connectionist networks) has been challenged in a number of linguistic domains. With respect to lexical access, Luce and Lyons (1998) provided new evidence from priming tasks for both abstract lexical units and stimulus-specific exemplars which can be accessed under different processing conditions. At the phonological level, Marcus and his colleagues (e.g., Berent, Marcus, Shimron, & Gafos, 2002) argued that humans must have abstract linguistic variables in order to account for their ability to generalize to novel stimuli that lie beyond the scope of actual linguistic experience, such as novel phoneme categories. Similarly, Pierrehumbert's models of phonological change (2001, 2002) and acquisition (2003) include multiple levels of linguistic representation that have both abstract and exemplar representations. She argued that exemplar traces of individual items help in accounting for lexical frequency effects, phonological shifts, and the development of robust phonological categories, but that abstract units are also necessary to explain the presence of systematic phonological processes that occur in human languages such as allophonic variation and phonotactic regularities. Dialect variation presents another domain of language that can be used to explore the role of both abstract units and exemplar-based encoding of linguistic information.

Pierrehumbert (2001, 2002) developed a model of the lexicon which contains representations at lexical, phonological, and phonetic levels. In her model, individual words are linked to their constituent phonemes which, in turn, are linked to phonetic elements. These different levels of linguistic representation are reminiscent of more modular accounts of linguistic systems in which strings from one level are manipulated in some way and then passed on to the next lower level. However, linguistic experience is also relevant at each level in her model. Lexical frequency, regular phonological change, and non-native phonetic instantiations can each be represented and accounted for under the exemplar approach.

Pierrehumbert's (2001) model of phonological change is instantiated using a category label *L* and a list of exemplars $\{e_1^L, e_2^L, \ldots, e_n^L\}$ associated with that category. Perception is modeled by computing goodness scores of candidate categories and assigning the stimulus to the category with the highest score. Production is modeled by randomly selecting one exemplar from the category *L* to serve as the target production, based on relative probabilities of the exemplars within the category. Learning is achieved through the interaction of perception and production processes.

Pierrehumbert's model could be extended to incorporate dialect variation by the addition of several other levels of representation. Figure 1 depicts a schematic of Pierrehumbert's (2001, 2002) model, with levels of representation added for individual talkers, dialects, and other social categories, such as friendliness and intelligence, which have been shown to be related to the perception of dialect variation (Ryan & Giles, 1982). At the dialect level, frequency and experience are related to the robust development of perceptual categories. Listeners are able to reliably perceive and produce their own native dialect. The ability to process and identify other dialects is limited by direct experience with that particular variety. Those features which are perceptually salient are more likely to be encoded and therefore more likely to be available and used in a wide range of behavioral tasks such as categorization and classification (Pierrehumbert, 2002).

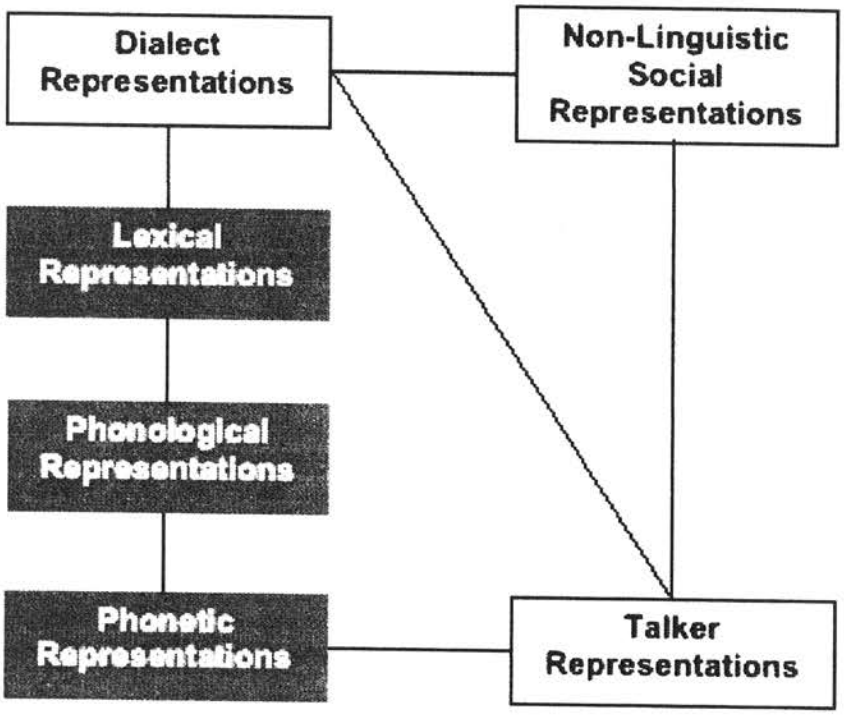


Figure 1. Schematic of an exemplar-based model of sociolinguistic variation. Pierrehumbert's (2001, 2002) original model is shown in gray. The new components added to account for sociolinguistic variation are shown in white.

Dialect category representations are linked to the phonological model to create a situation in which multiple levels of representation allow for overlapping categories at the phonetic level which are in turn linked to distinct categories at the phonological, lexical, and dialect levels. Figure 2 provides an example of how the model might account for phonetic variation, such as the difference between the vowel systems of Northern and Southern American English talkers. Recall that Northern talkers produce a fronted and raised [æ] as well as a backed [ɛ] as part of the Northern Cities Chain Shift. Southern talkers, on the other hand, produce unshifted [æ]s and raised [ɛ]s. The result is that some Northern [æ]s and some Southern [ɛ]s are produced in overlapping areas of the vowel space. However, the differences between the relevant phonetic and phonological information allow the listener to correctly process the phonemic and lexical information and provide an indication about the dialect of the talker. For example, if the phonetic instantiation is somewhere near an [ɛ] but the phoneme category is /æ/, it is likely that the talker comes from the Northern dialect region.

Representations of the dialect categories are also linked to specific talkers as well as other, non-linguistic attributes that are encoded in the speech signal. Additional research is needed to explore how these three sets of representations are acquired. The model presented in Figures 1 and 2 also assumes that speech intelligibility is perfect, even under conditions where there is a great deal of variability due to differences in the dialects of the talkers and the listeners. However, Mason's (1946) study suggested that this is not always the case. Therefore, additional modeling may be necessary to account for cases of misperception at the phonological and lexical level as it relates to dialect variation.

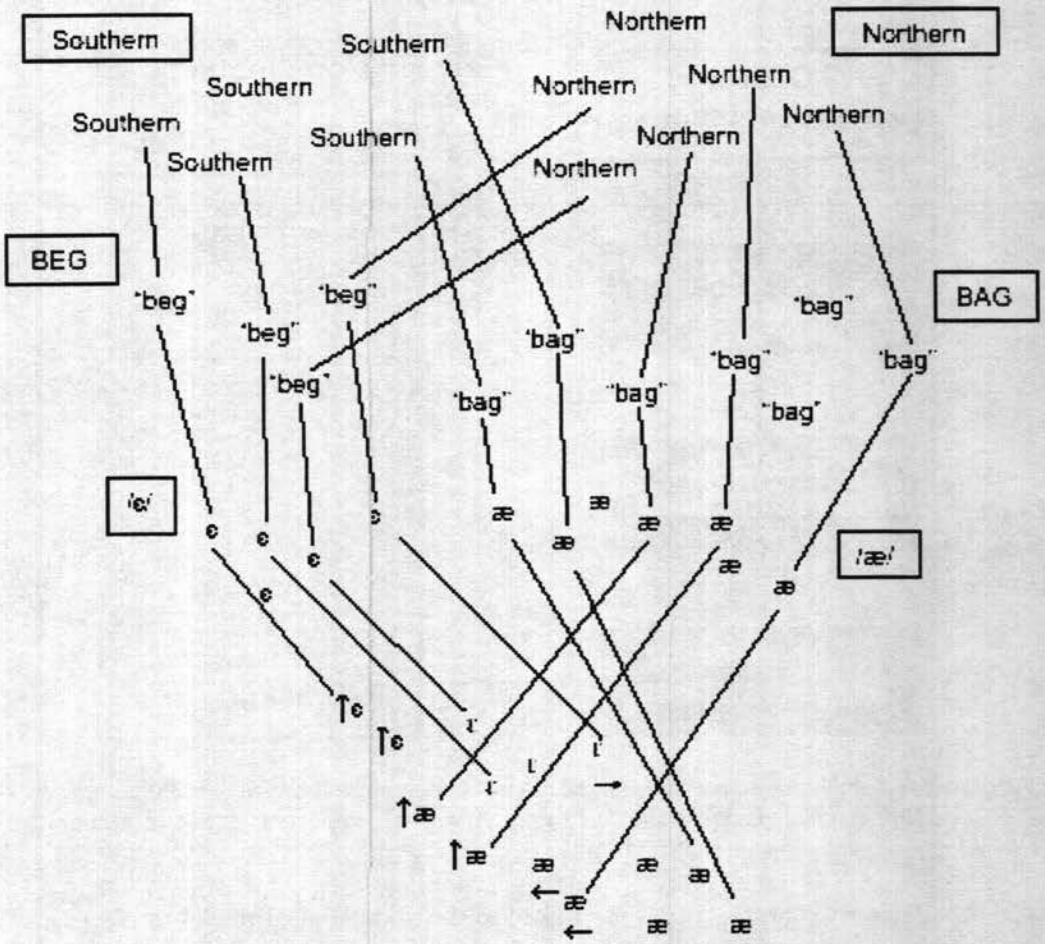


Figure 2. Example of the relationship between phonetic, phonological, lexical, and dialect representations in an exemplar-based model of sociolinguistic variation.

Conclusions

The research presented in this dissertation extends the earlier work carried out by Clopper and colleagues (Clopper et al., in press; Clopper & Pisoni, 2004a, b) on the perception of dialect variation in several new and important ways. First, we collected a new speech corpus that was explicitly designed for use in acoustic analyses and perceptual categorization studies of regional variation in the United States. Second, several issues related to the residential history of the listeners were examined in more detail and significant effects of both geographic location and mobility were found with respect to the underlying perceptual similarity structure of regional variation for naïve listeners. Third, a free classification task was used to measure the perceptual similarity of the dialects of the talkers without the constraints of specific verbal labels or a closed set of responses structured by specific geographic regions. The results of these new experiments were consistent with the previous forced-choice results, but also revealed several apparent response biases in the forced-choice task. Finally, the findings were discussed in terms of several broader theoretical issues in linguistics and cognitive science, specifically the role of perception in sociolinguistics, the role of variation in speech perception and spoken language processing, and the plausibility of an exemplar-based model of phonological systems which could account for social variation along with a range of other linguistically significant properties of spoken language.

References

Bennett, S., & Montero-Diaz, L. (1982). Children's perception of speaker sex. *Journal of Phonetics, 10*, 113-121.

Berent, I., Marcus, G.F., Shimron, J., & Gafos, A.I. (2002). The scope of linguistic generalizations: Evidence from Hebrew word formation. *Cognition, 83*, 113-139.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Carver, C.M. (1987). *American Regional Dialects: A Word Geography*. Ann Arbor, MI: University of Michigan Press.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clopper, C.G., Conrey, B.L., & Pisoni, D.B. (in press). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology*.

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.

Goldinger, S.D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166-1183.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J.W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-165). San Diego: Academic Press.

Joos, M. (1948). Acoustic phonetics. *Language, 24*, 5-136.

Jusczyk, P.W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

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.

Krapp, G.P. (1925). *The English Language in America*. New York: Frederick Ungar.

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 dialects of English. In M.D. Linn (Ed.), *Handbook of Dialects and Language Variation* (pp. 39-81). San Diego: Academic Press.

Labov, W., Ash, S., & Boberg, C. (forthcoming). *Atlas of North American English*. New York: Mouton de Gruyter.

Lass, N.J., Hughes, K.R., Bowyer, M.D., Waters, L.T., & Bourne, V.T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America, 59*, 675-678.

Luce, P.A., & Lyons, E.A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition, 26*, 708-715.

Mason, H. (1946). Understandability of speech in noise as affected by region of origin of speaker and listener. *Speech Monographs, 13*, 54-58.

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.

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

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

Phillips, B.S. (2001). Lexical diffusion, lexical frequency, and lexical analysis. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 123-136). Amsterdam: John Benjamins.

Pierrehumbert, J.B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137-157). Amsterdam: John Benjamins.

Pierrehumbert, J.B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 101-140). Berlin: Mouton de Gruyter.

Pierrehumbert, J.B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115-154.

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.

Plichta, B., & Preston, D.R. (2003). The /ay/s have it: Stereotype, perception, and region. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

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

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

Rakerd, B., & Plichta, B. (2003). More on perceptions of /ɑ/ fronting. Paper presented at New Ways of Analyzing Variation 32, Philadelphia, PA. October 9-12.

Remez, R.E., Fellowes, J.M., & Rubin, P.E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 651-666.

Ryan, E.B., & Giles, H. (1982). *Attitudes towards Language Variation.* London: Edward Arnold.

Sheffert, S.M., Pisoni, D.B., Fellowes, J.M., & Remez, R.E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1447-1469.

Stevens, K.N. (1996). Understanding variability in speech: A requisite for advances in speech synthesis and recognition. *Journal of the Acoustical Society of America, 100*, 2634.

APPENDIX A: THE NATIONWIDE SPEECH PROJECT CORPUS MATERIALS

hVd Words

hoed
heed
hid
hayed
head
had
hod
hud
hood
who'd

CVC Words

bean	bite	boat	boil
bull	cab	calm	can
caught	coal	code	coin
con	cool	cot	cough
death	dig	dime	dock
doll	doubt	dull	fade
fail	feed	fell	fire
fool	foul	full	gap
good	guide	head	heal
hill	home	keep	lit
loud	love	loyal	luck
lull	lung	main	math
meal	mile	mill	mob
pal	pen	pin	poke
pool	pull	rice	rip
sail	sell	sour	south
tape	tool	towel	town
tube	voice	void	walk
wall	wet	wool	wrong

Multisyllabic Words

absent	absentee	accelerate	accuser
adapt	alibi	alien	alligator
amphibian	anonymous	aquarium	armadillo
athletic	axle	bassinet	bazooka
bikini	bombard	café	caterpillar
cauliflower	cemetery	circumstance	clarinet
constellation	consume	coronation	coyote
dandelion	deactivate	detrimental	diabetes
disagree	disapproval	discourteous	earthy
equalizer	evaporate	exterior	feminine
feminism	forecast	functionary	gallop

generalize	giant	guitar	happily
helicopter	hurricane	incoherent	influenza
insulate	iodine	kangaroo	kazoo
lemonade	liberator	macaroni	malfunction
manufacture	martini	mascara	mighty
mispronounce	museum	nectarine	obedient
obsession	obtain	opportune	optimism
outdoors	outshine	paragraph	paranoia
pastel	peninsula	photographic	poisoner
porcelain	porcupine	potassium	precipitate
preferably	proclaim	procrastinate	prolong
qualify	readable	recording	robin
rodeo	salamander	sarcastic	serenade
spider	sterile	strengthen	subdivide
superhighway	sweeten	tangerine	teamster
tiger	tomato	trombone	unadvised
unbroken	unclean	underrate	victorious

High Probability Sentences

A bicycle has two wheels.
A spoiled child is a brat.
Banks keep their money in a vault.
Break the dry bread into crumbs.
Eve was made from Adam's rib.
For dessert he had apple pie.
Greet the heroes with loud cheers.
He was scared out of his wits.
Her hair was tied with a blue bow.
I ate a piece of chocolate fudge.
It was stuck together with glue.
Keep your broken arm in a sling.
Maple syrup is made from sap.
My son has a dog for a pet.
Our seats were in the second row.
Paul took a bath in the tub.
Peter dropped in for a brief chat.
Please wipe your feet on the mat.
Ruth had a necklace of glass beads.
She cooked him a hearty meal.
Spread some butter on your bread.
The bird of peace is the dove.
The boat sailed across the bay.
The burglar escaped with the loot.
The candle burned with a bright flame.
The car was parked at the curb.
The chicks followed the mother hen.
The cow gave birth to a calf.
The dealer shuffled the cards.
The doctor prescribed the drug.
A round hole won't take a square peg.
Ann works in the bank as a clerk.
Bob was cut by the jacknife's blade.
Cut the meat into small chunks.
Follow this road around the bend.
Get the bread and cut me a slice.
He rode off in a cloud of dust.
Her entry should win first prize.
He's employed by a large firm.
Instead of a fence, plant a hedge.
I've got a cold and a sore throat.
Kill the bugs with this spray.
My jaw aches when I chew gum.
Old metal cans were made with tin.
Paul hit the water with a splash.
Paul was arrested by the cops.
Playing checkers can be fun.
Raise the flag up the pole.
Ruth poured herself a cup of tea.
She shortened the hem of her skirt.
That job was an easy task.
The bloodhound followed the trail.
The bride wore a white gown.
The cabin was made of logs.
The car drove off the steep cliff.
The chicken pecked corn with its beak.
The cigarette smoke filled his lungs.
The cut on his knee formed a scab.
The detectives searched for a clue.
The firemen heard her frightened scream.

The flashlight casts a bright beam.
The gambler lost the bet.
The guests were welcomed by the host.
The house was robbed by a thief.
The landlord raised the rent.
The mouse was caught in the trap.
The plow was pulled by an ox.
The poor man was deeply in debt.
The shepherd watched his flock of sheep.
The sick child swallowed the pill.
The story had a clever plot.
The swimmer dove into the pool.
The thread was wound on a spool.
The witness took a solemn oath.
They tracked the lion to his den.
Throw out all this useless junk.
To open the jar, twist the lid.
Tree trunks are covered with bark.
Wash the floor with a mop.
We camped out in our tent.
We swam at the beach at high tide.
The flood took a heavy toll.
The glass had a chip on the rim.
The heavy rains caused a flood.
The judge is sitting on the bench.
The lion gave an angry roar.
The nurse gave him first aid.
The pond was full of croaking frogs.
The scarf was made of shiny silk.
The shepherds guarded their flock.
The stale bread was covered with mold.
The super highway has six lanes.
The swimmer's leg got a bad cramp.
The wedding banquet was a feast.
They drank a whole bottle of gin.
This camera is out of film.
Tighten the belt by a notch.
To store his wood, he built a shed.
Unlock the door and turn the knob.
Watermelons have lots of seeds.
We heard the ticking of the clock.
Your knees and your elbows are joints.

Low Probability Sentences

Mr. Smith spoke about the aid.
Mr. Smith knew about the bay.
She's glad Bill called about the beak.
I'm talking about the bench.
Mary hasn't discussed the blade.
Mr. Black has discussed the cards.
Bob was considering the clerk.
Ruth hopes Bill called about the cop.
We will consider the debt.
We could consider the feast.
Paul should have discussed the flock.
The woman talked about the frogs.
Tom has not considered the glue.
Bill didn't discuss the hen.
Harry had thought about the logs.
Peter has considered the mat.
He has a problem with the oath.
They heard I called about the pet.
Tom had spoken about the pill.
She might consider the pool.
You've considered the seeds.
Tom won't consider the silk.
Bob has discussed the splash.
Bob considered the tent.
We're speaking about the toll.
Mr. Brown thinks about the vault.
Betty has considered the bark.
Tom has been discussing the beads.
She's discussing the beam.
I'm glad you heard about the bend.
She hopes Jane called about the calf.
I did not know about the chunks.
The man spoke about the clue.
She might discuss the crumbs.
Peter could consider the dove.
The girl should consider the flame.
The class should consider the flood.
The girl talked about the gin.
The girl should not discuss the gown.
Bill heard we asked about the host.
The old man talked about the lungs.
The woman considered the notch.
The man should discuss the ox.
Ruth must have known about the pie.
Bob could consider the pole.
The boy would discuss the scab.
They've considered the sheep.
Nancy didn't discuss the skirt.
Mary had considered the spray.
Mary can't consider the tide.
Miss Smith knows about the tub.
Bill can't have considered the wheels.

Anomalous Sentences

The bread gave hockey loud aid.
The problem hoped under the bay.
The cat is digging bread on its beak.
The arm is riding on the bench.
Miss Smith was worn by Adam's blade.
The turn twisted the cards.
Jane ate in the glass for a clerk.
Nancy was poured by the cops.
Mr. White hit the debt.
The first man heard a feast.
The problems guessed their flock.
The coat is talking about six frogs.
It was beaten around with glue.
The stories covered the glass hen.
The ship was interested in logs.
Please throw out your firemen in the mat.
The thread broke a clenched oath.
My man wiped a shirt for a pet.
The folding hands drowned the pill.
The jar swept up the pool.
Heroes called lots of seeds.
The breakfast was decided by whole silk.
Tom took the elbow after a splash.
We rode off in our tent.
The king shipped a metal toll.
Highways pour their hair about a vault.
Ruth's problems are made from bark.
Bill knew a can of maple beads.
The accident washes a short beam.
Discuss this sailboat on the bend.
The low woman was gladly in the calf.
Toss the boy into shipwrecked chunks.
The landlords stood for a clue.
Consider the local floor about crumbs.
The player of maple is the dove.
The sink served with an easy flame.
The round lion held a flood.
They milked a frightened entry of gin.
Betty buttered a sharp gown.
The seats were called about the host.
The old cloud broke his lungs.
Face the cop through a notch.
The burglar was parked by an ox.
For a bloodhound he had spoiled pie.
Water the worker between the pole.
The chimpanzee on his checkers wore a scab.
Miss Brown charged her wood of sheep.
She considered the floor of Jane's skirt.
Think about the boys with this spray.
We cook at the car at sore tide.
Eve raised a team in the tub.
David knows long wheels.

Rainbow Passage

When sunlight strikes the raindrops in the air, they act like a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Goldilocks Passage

Once upon a time there were three bears: a Daddy bear, a Mommy bear and a little baby bear. They lived in a cottage deep in the woods. One morning, Mommy bear had made some porridge for breakfast, but it was too hot to eat at once. 'Let's go for a walk while it cools down,' said Daddy bear. 'What a good idea!' exclaimed Mommy bear, and, with their bear coats and bear shoes on, they all set off for a short walk in the woods. That morning a little girl called Goldilocks was also walking in the woods. She was picking flowers and had wandered deeper in among the trees than her parents allowed her to go. After a while of being completely lost, she came into a clearing and saw the pretty little cottage. 'I wonder who lives there?' she thought to herself, and walked up to the door. When she knocked, there was no answer, so she pushed the door. It swung open, and she went in.

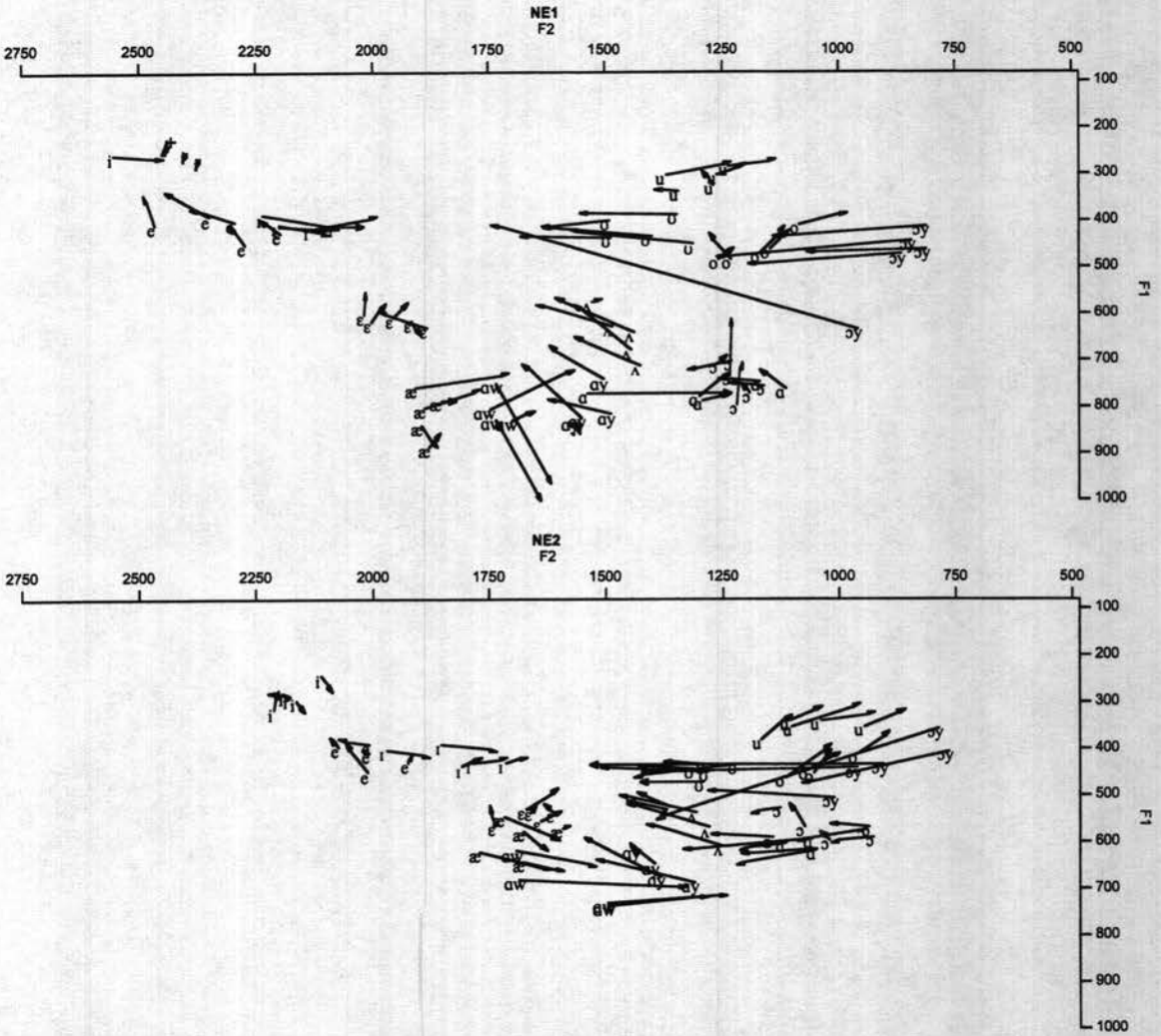
Targeted Interview Target Words

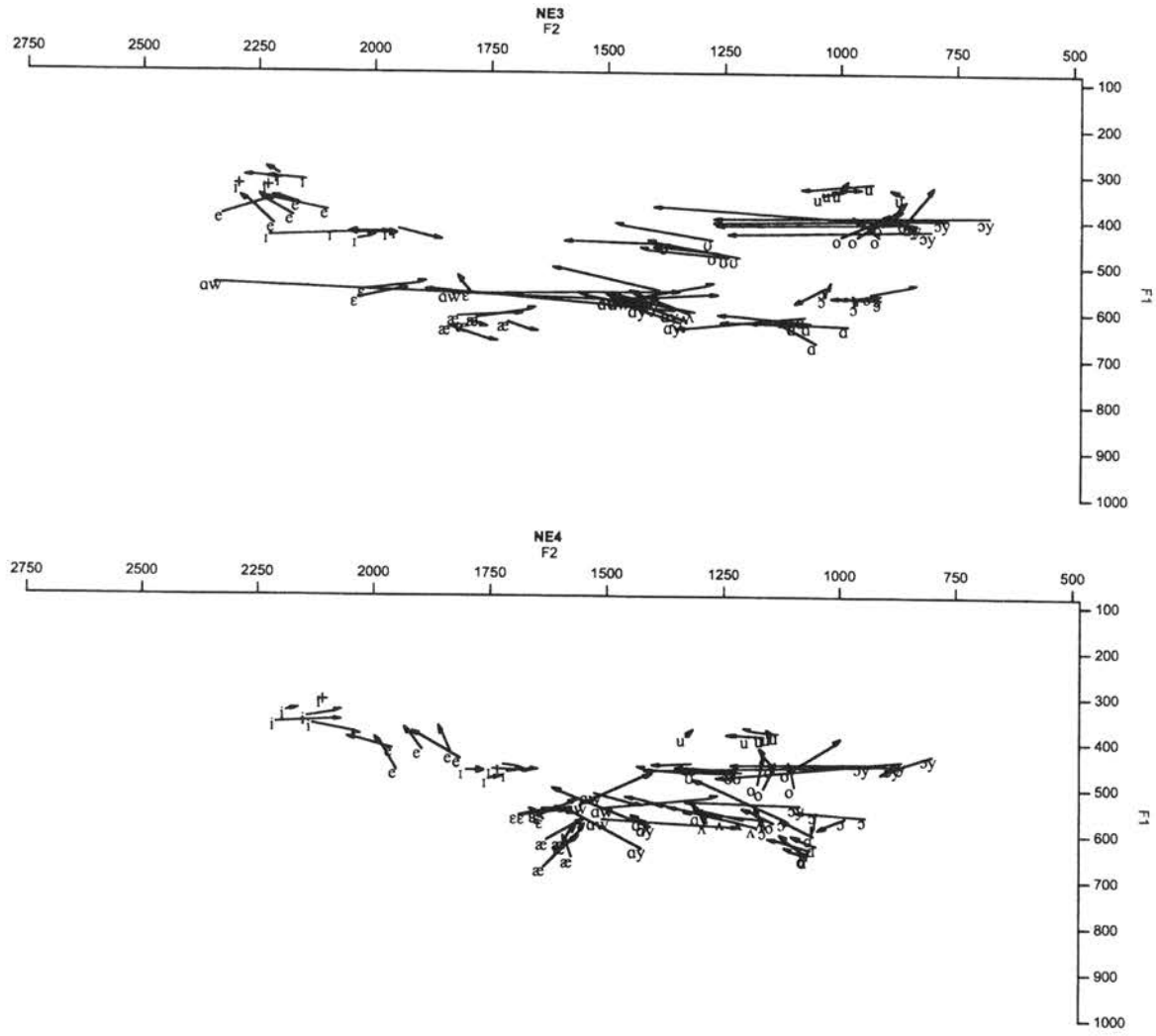
1. cook
2. fish
3. shoes
4. slip
5. date
6. luck
7. met
8. math
9. job
10. boat

APPENDIX B: THE NATIONWIDE SPEECH PROJECT CORPUS INDIVIDUAL VOWEL SPACES

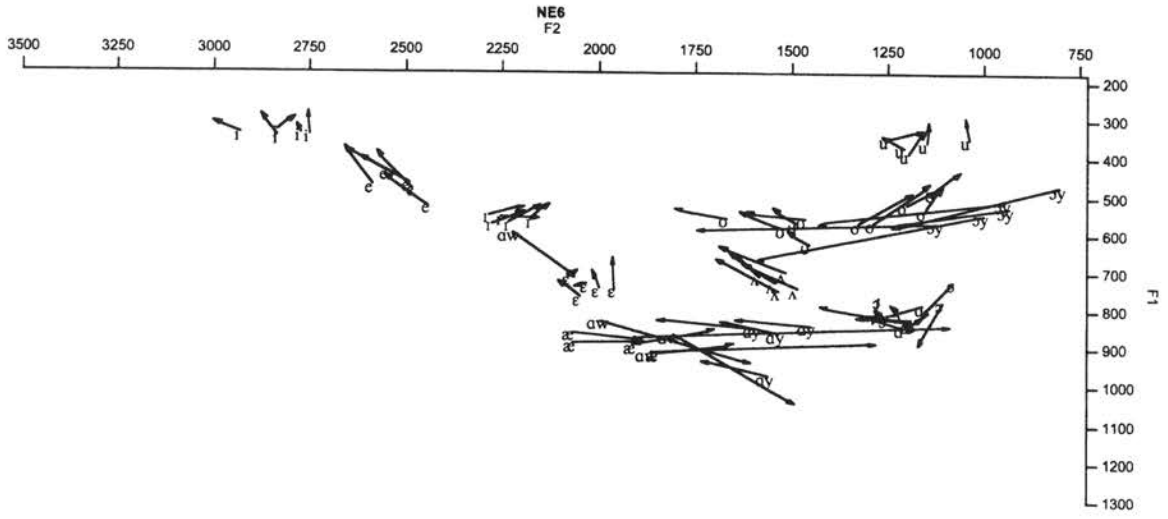
The figures in Appendix B are individual vowel spaces for each of the 48 talkers in the NSP corpus that were included in the acoustic analysis in Chapter 2. In each figure, all 69 vowel tokens for an individual talker were plotted in an F1 x F2 space. The phonetic symbols indicate the vowel category of the token. The symbols are located at the first and second formant frequency values obtained at the first-third temporal point of the vowel. The arrows indicate formant frequency change over time and the arrow heads are located at the first and second formant frequency values obtained at the second-third temporal point of the vowel.

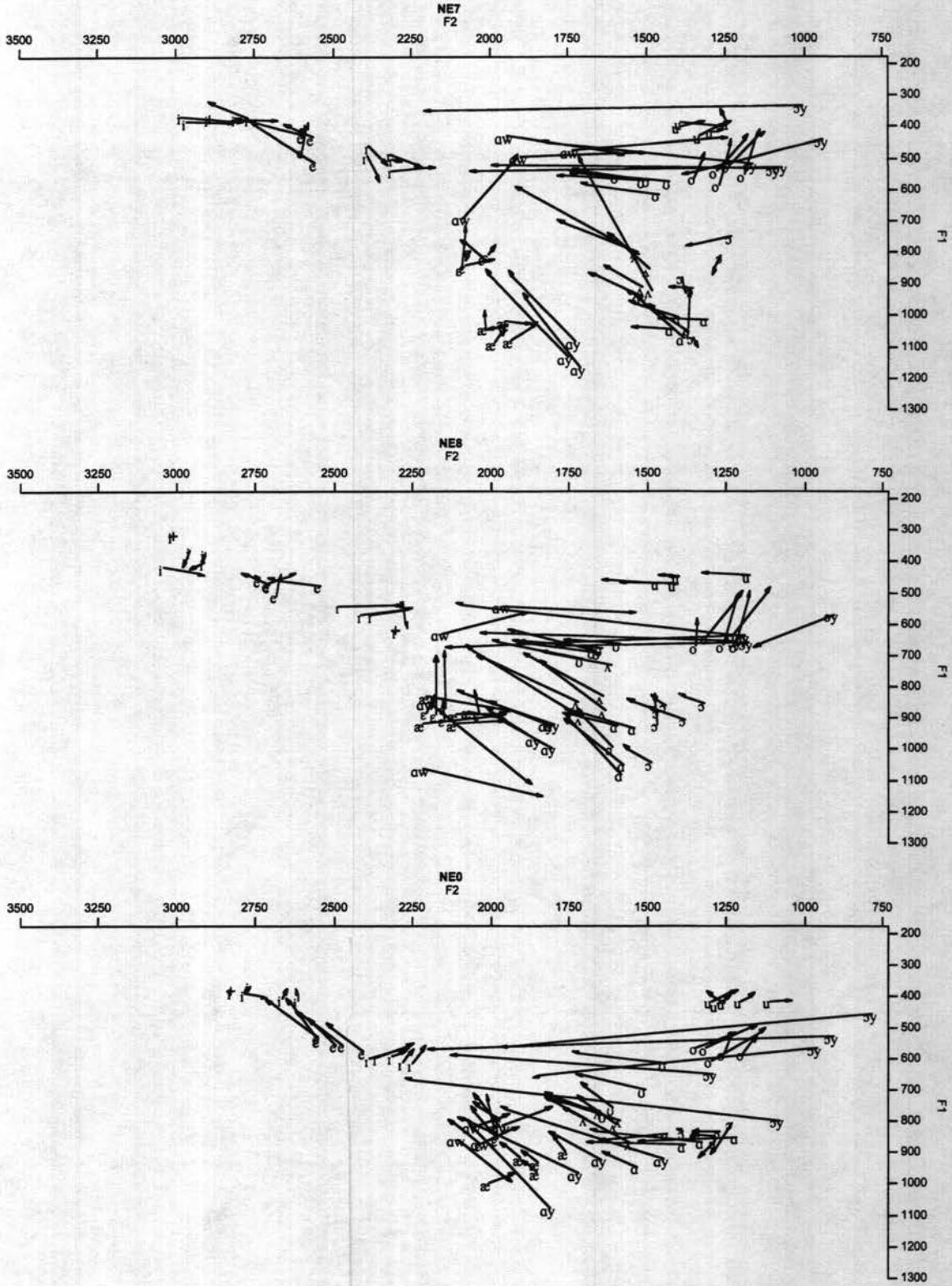
New England Males



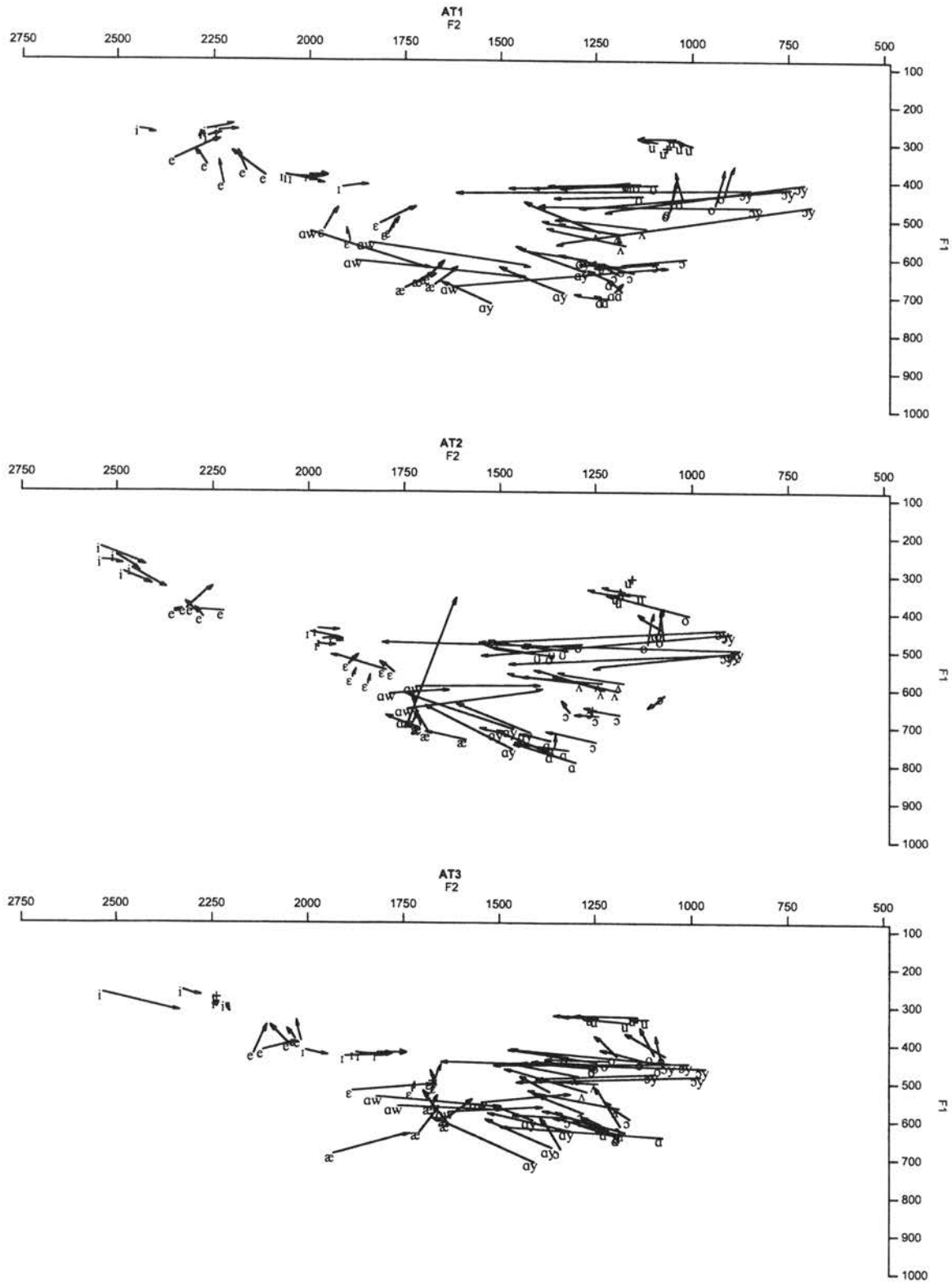


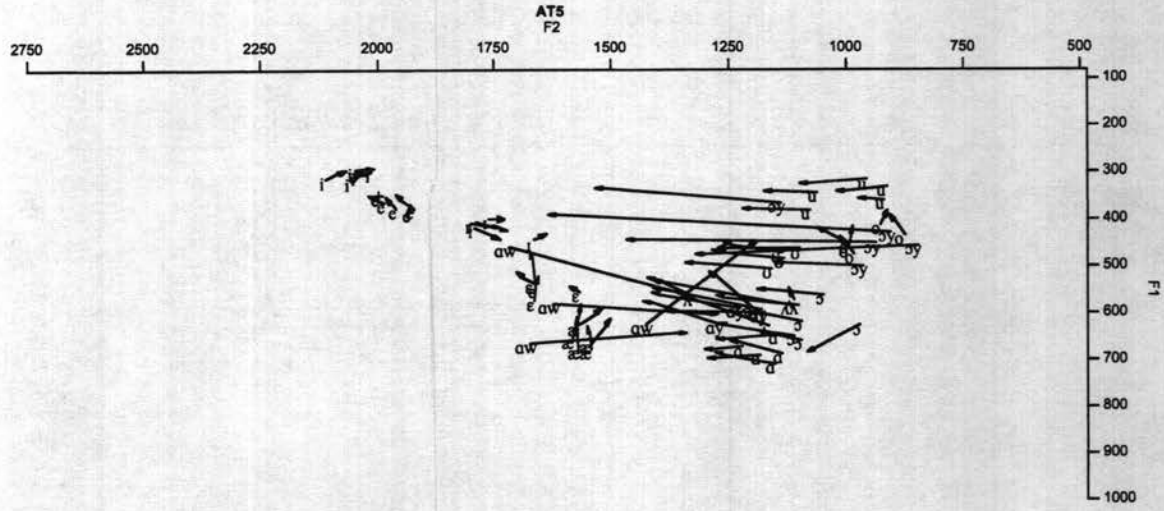
New England Females



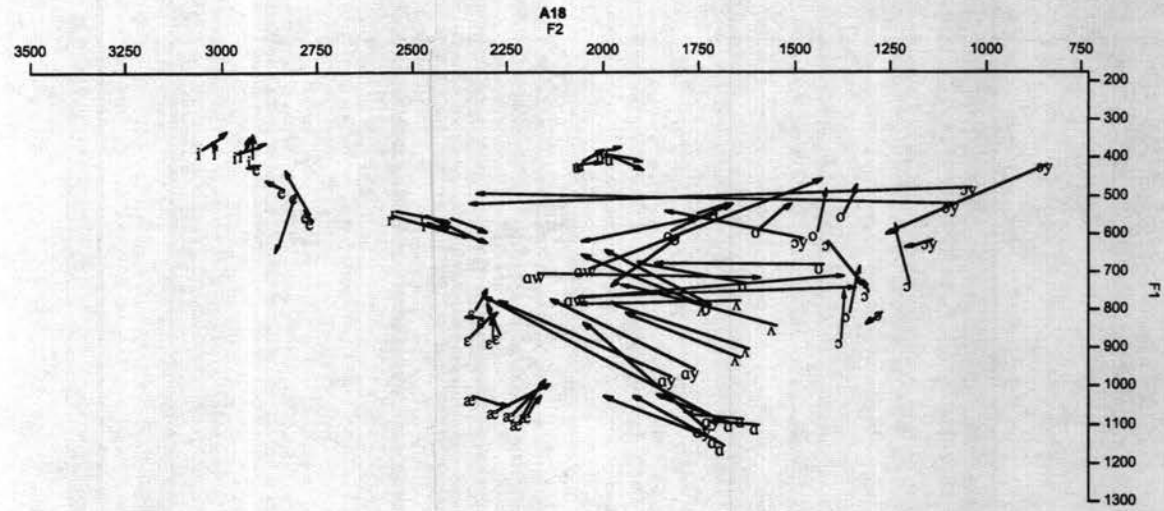

NE7
F2
3500 3250 3000 2750 2500 2250 2000 1750 1500 1250 1000 750
200 300 400 500 600 700 800 900 1000 1100 1200 1300
F1
NE8
F2
3500 3250 3000 2750 2500 2250 2000 1750 1500 1250 1000 750
200 300 400 500 600 700 800 900 1000 1100 1200 1300
F1
NE0
F2
3500 3250 3000 2750 2500 2250 2000 1750 1500 1250 1000 750
200 300 400 500 600 700 800 900 1000 1100 1200 1300
F1

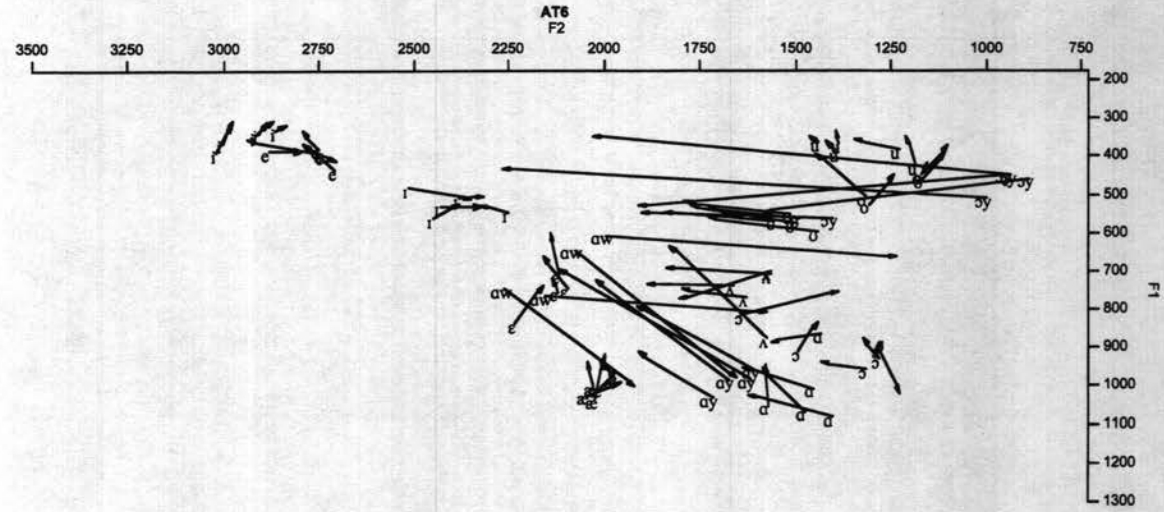
Mid-Atlantic Males

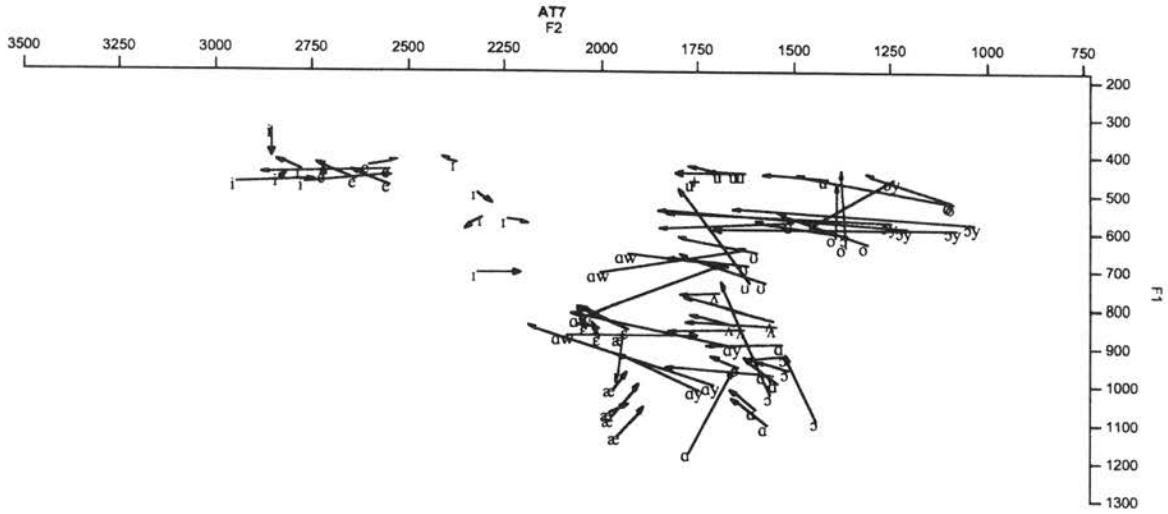


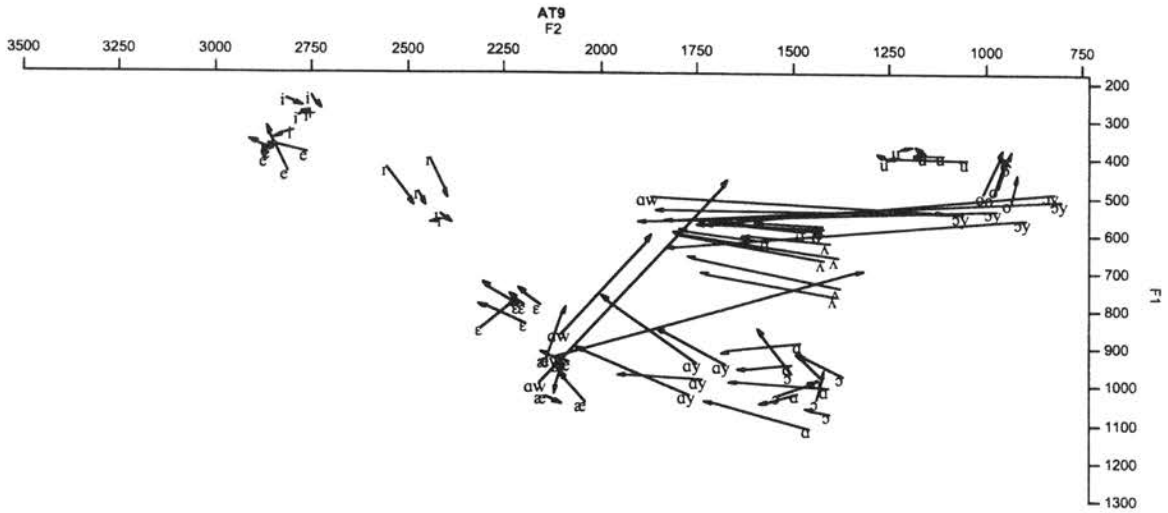


Mid-Atlantic Females

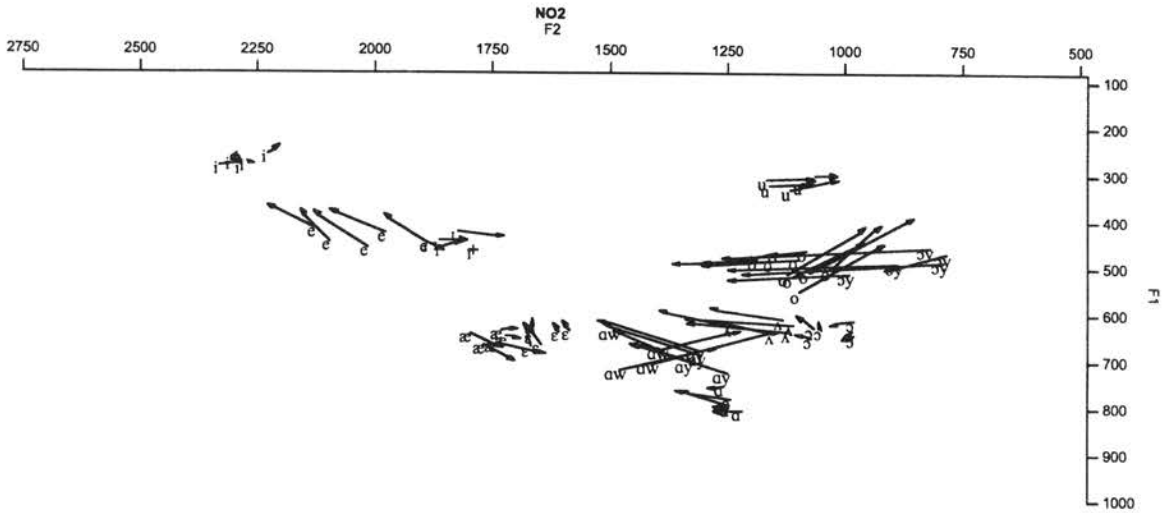


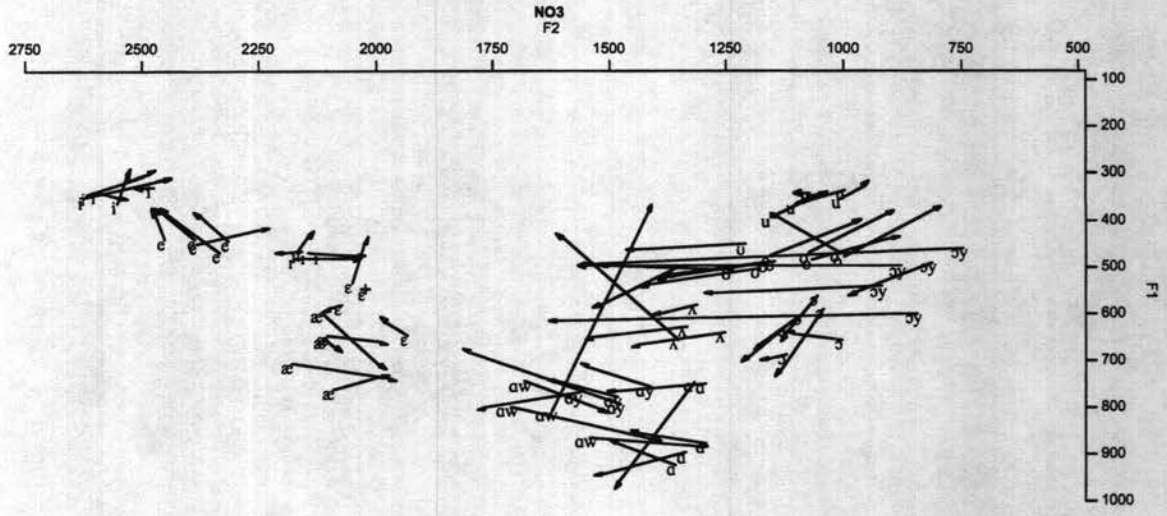


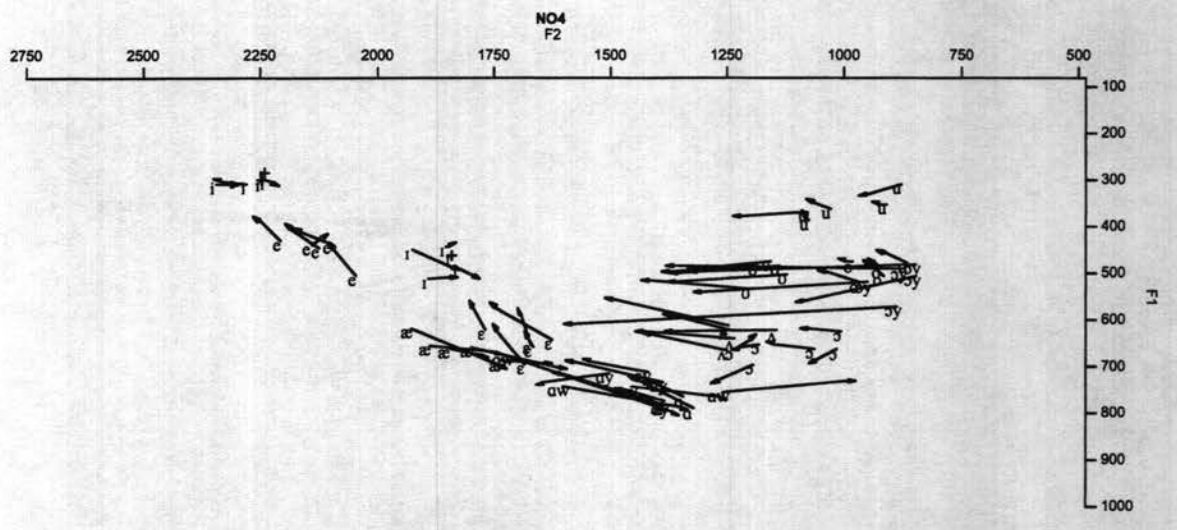


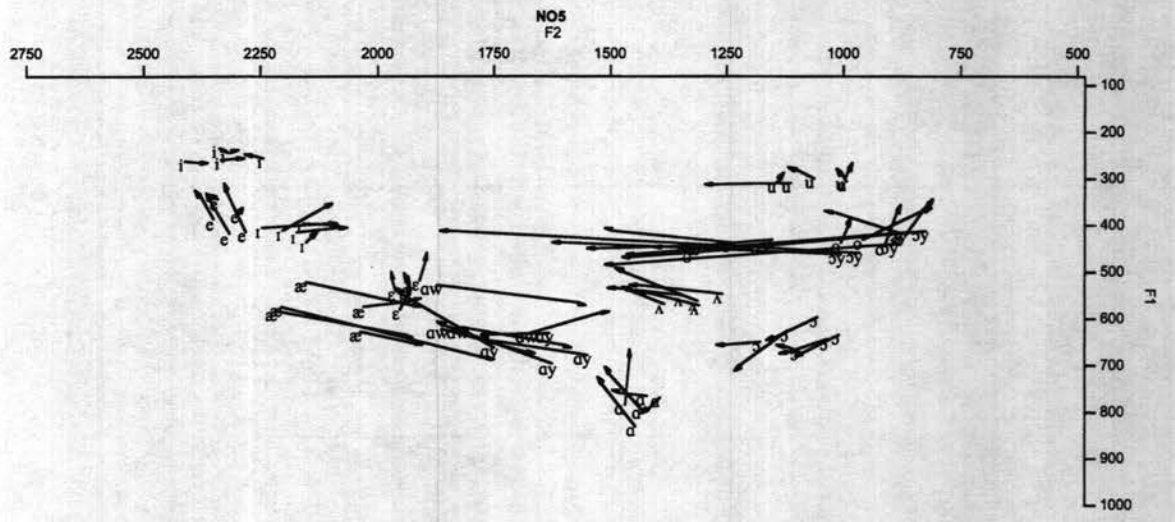


Northern Males

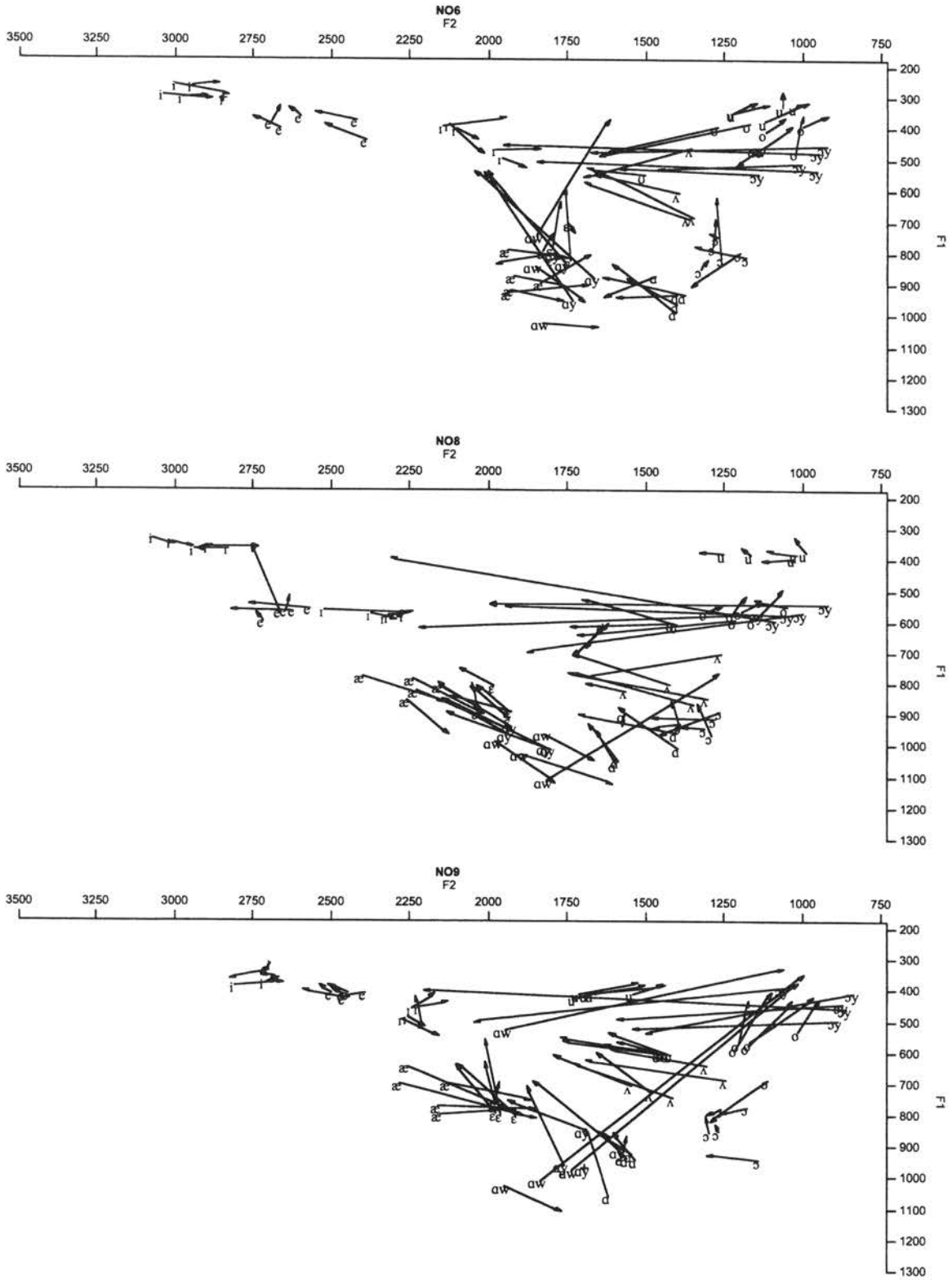


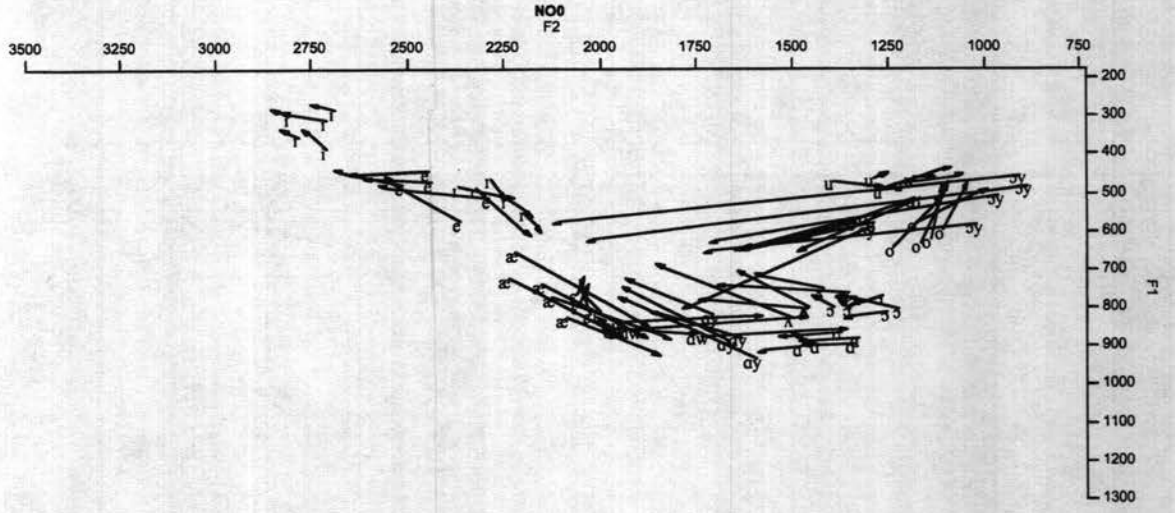

NO3
F2
2750
2500
2250
2000
1750
1500
1250
1000
750
500
100
200
300
400
500
600
700
800
900
1000
F1


NO4
F2
2750
2500
2250
2000
1750
1500
1250
1000
750
500
100
200
300
400
500
600
700
800
900
1000
F1

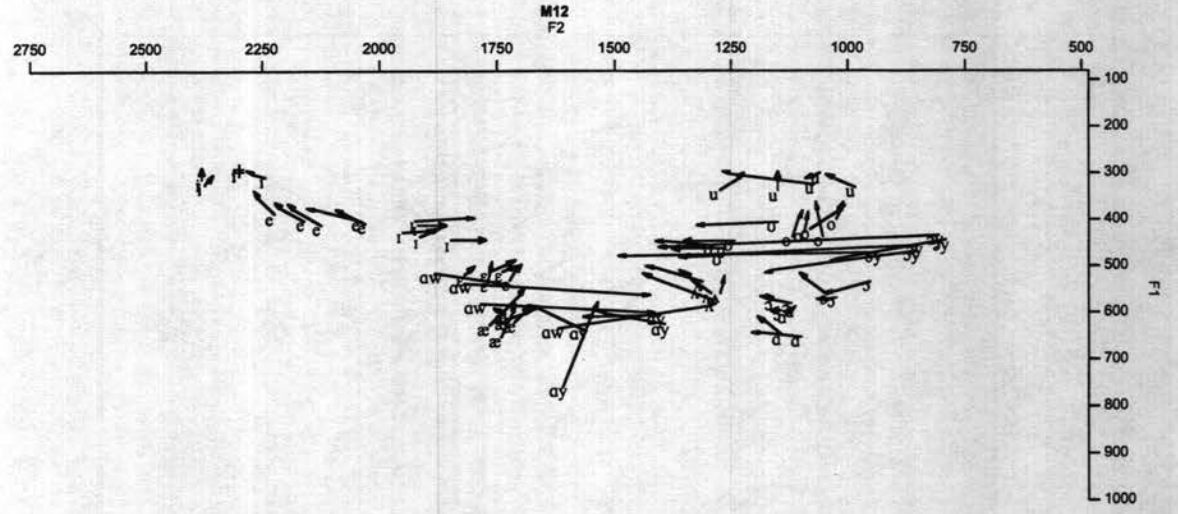

NO5
F2
2750
2500
2250
2000
1750
1500
1250
1000
750
500
100
200
300
400
500
600
700
800
900
1000
F1

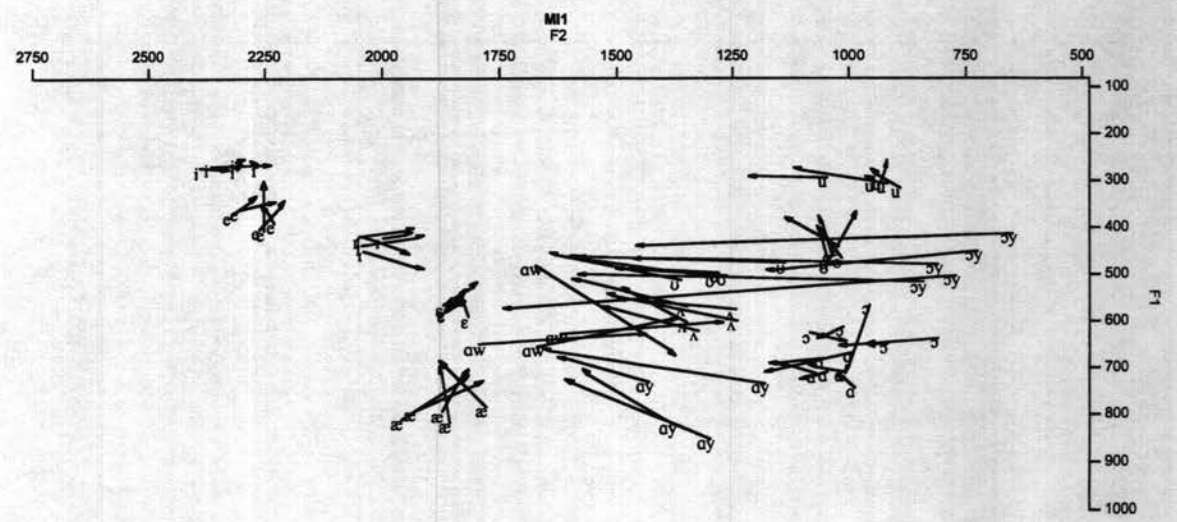
Northern Females

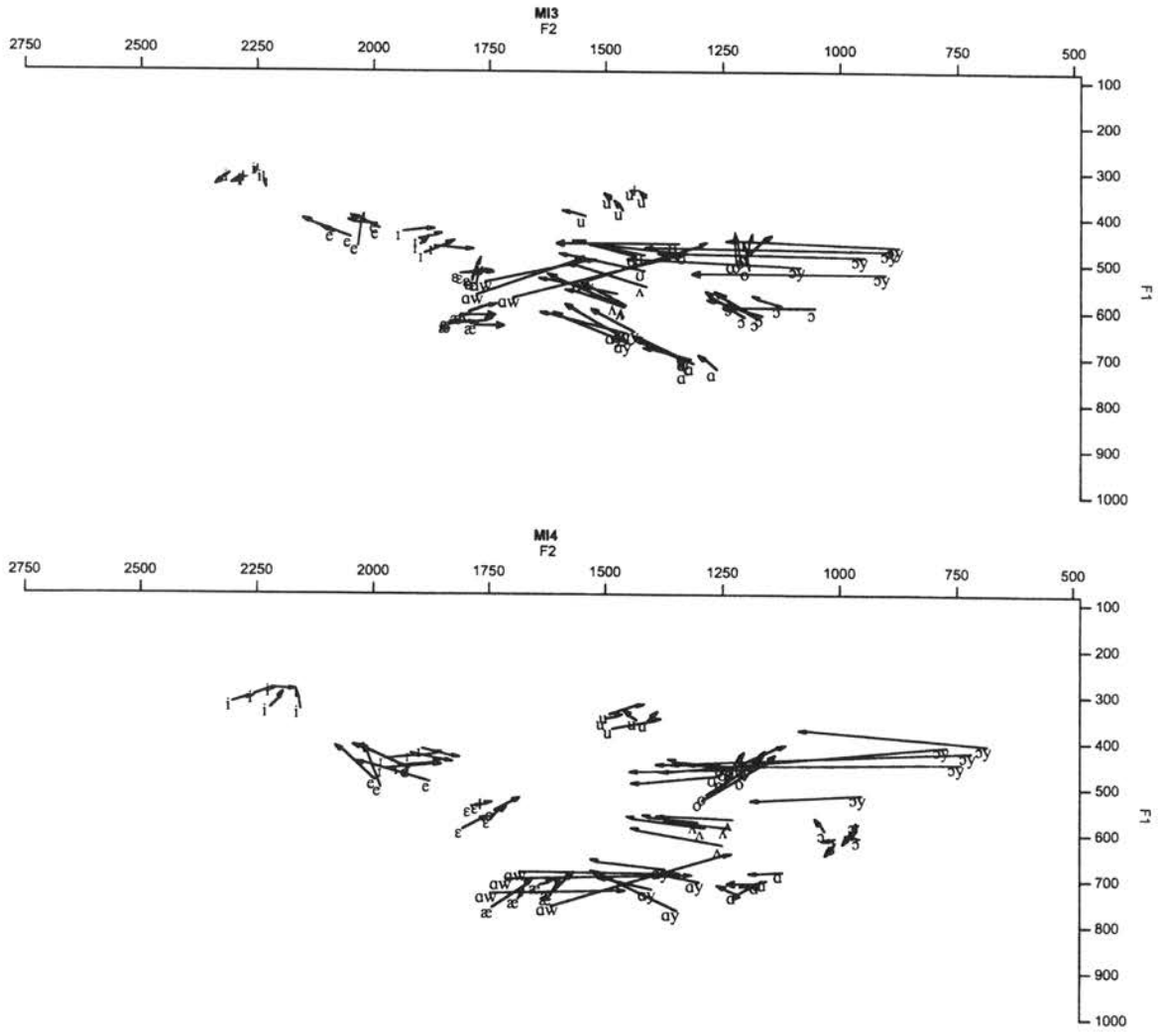




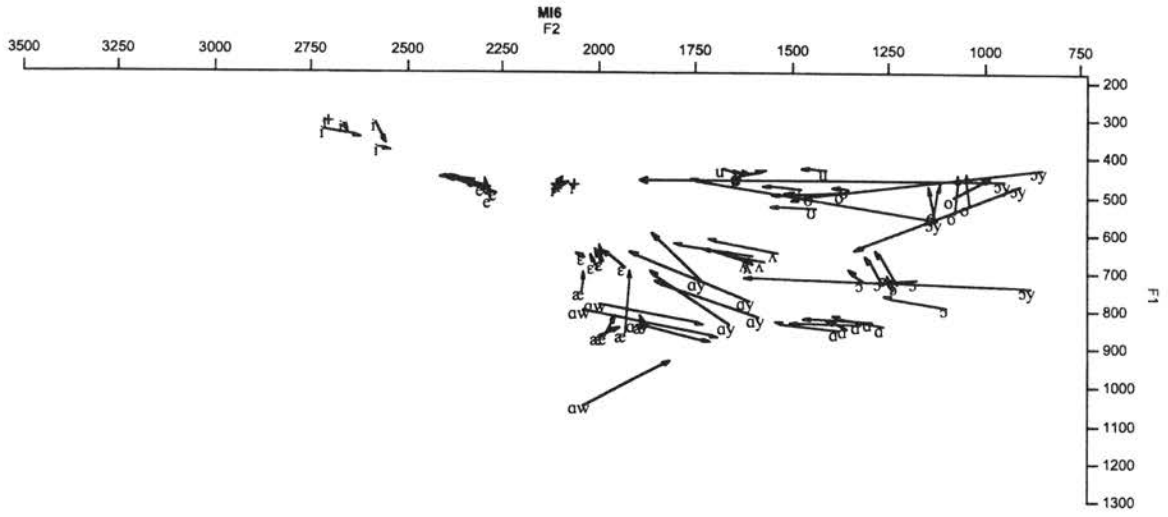
Midland Males

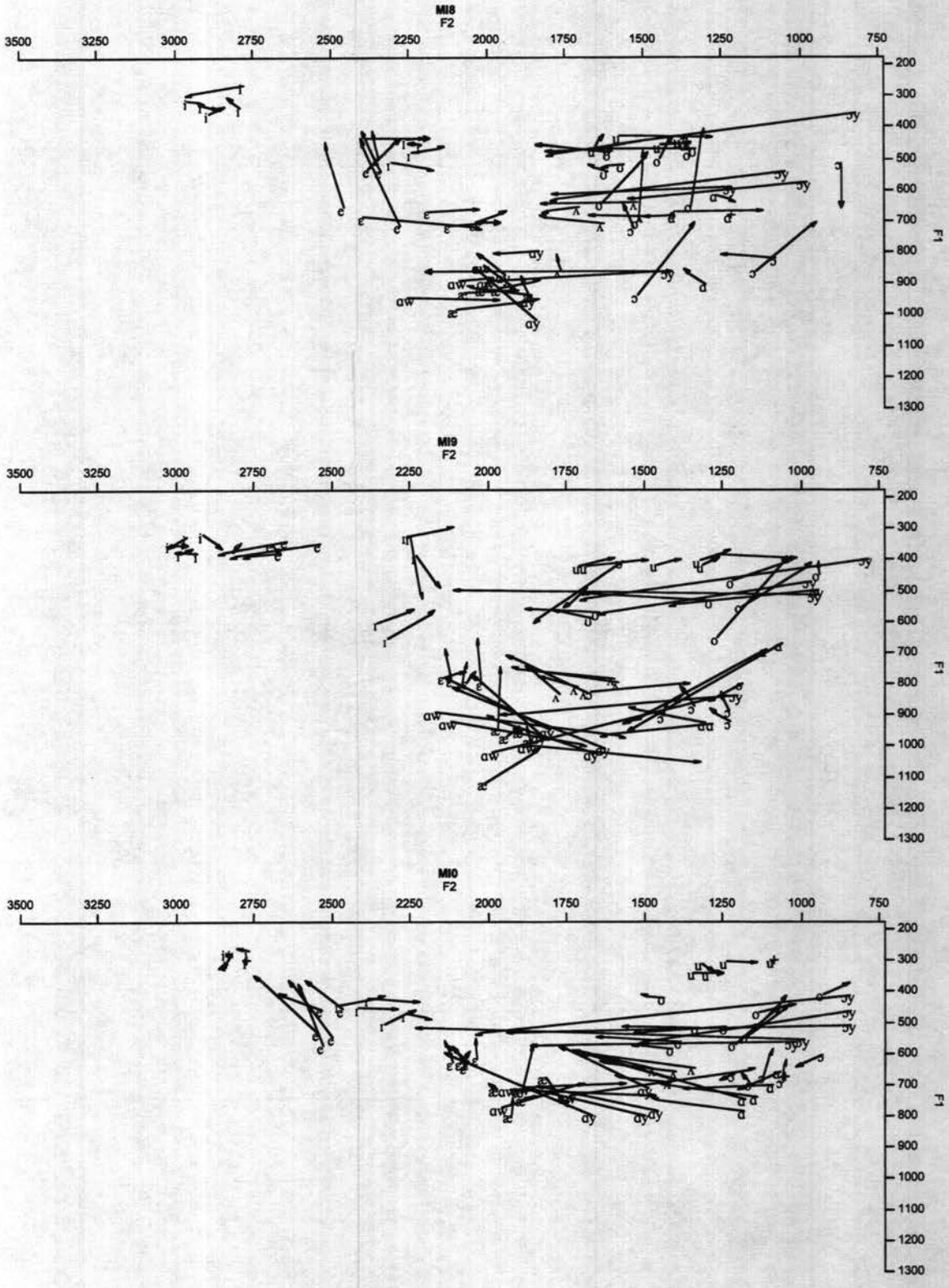




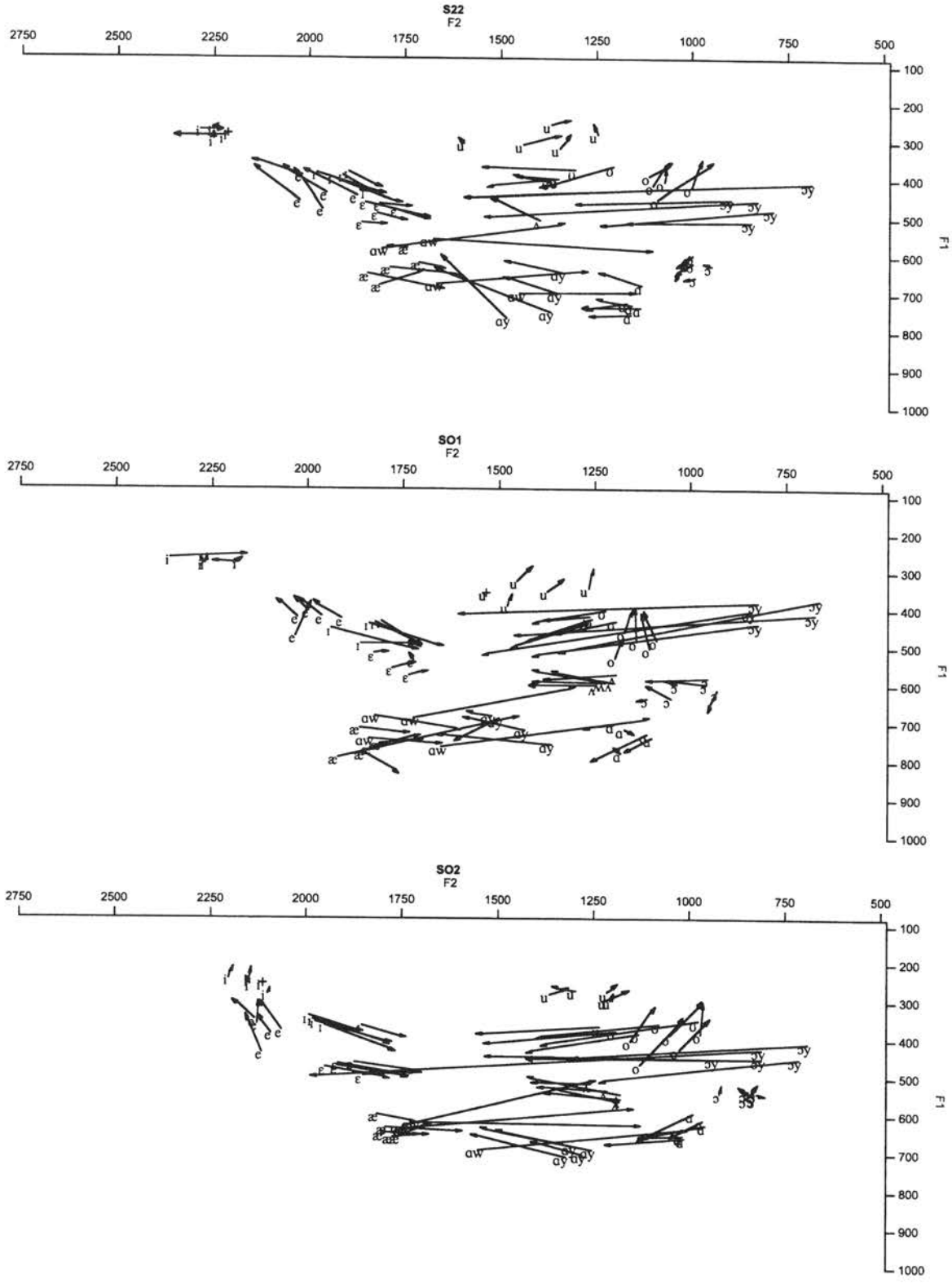


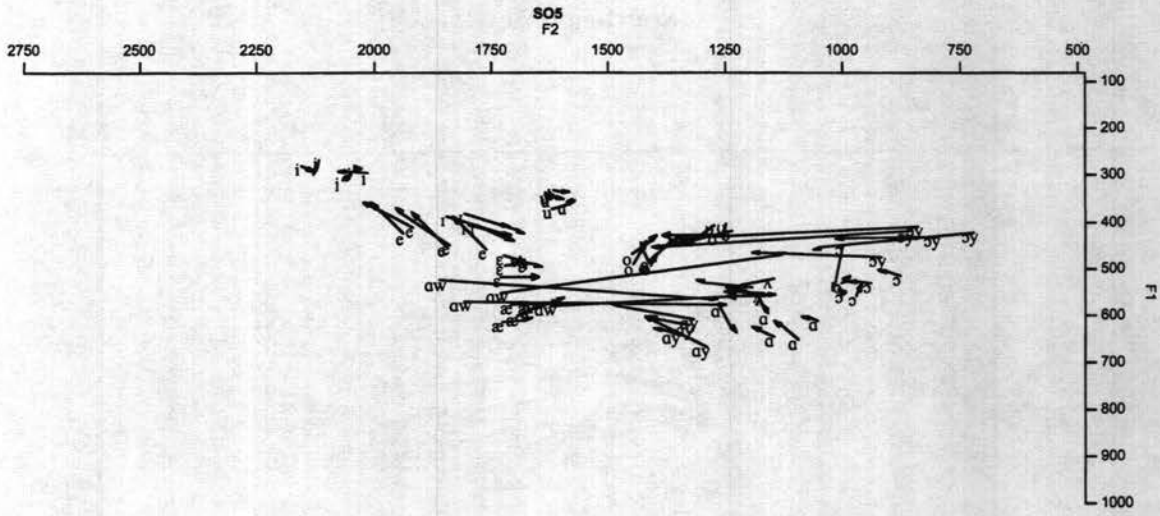
Midland Females



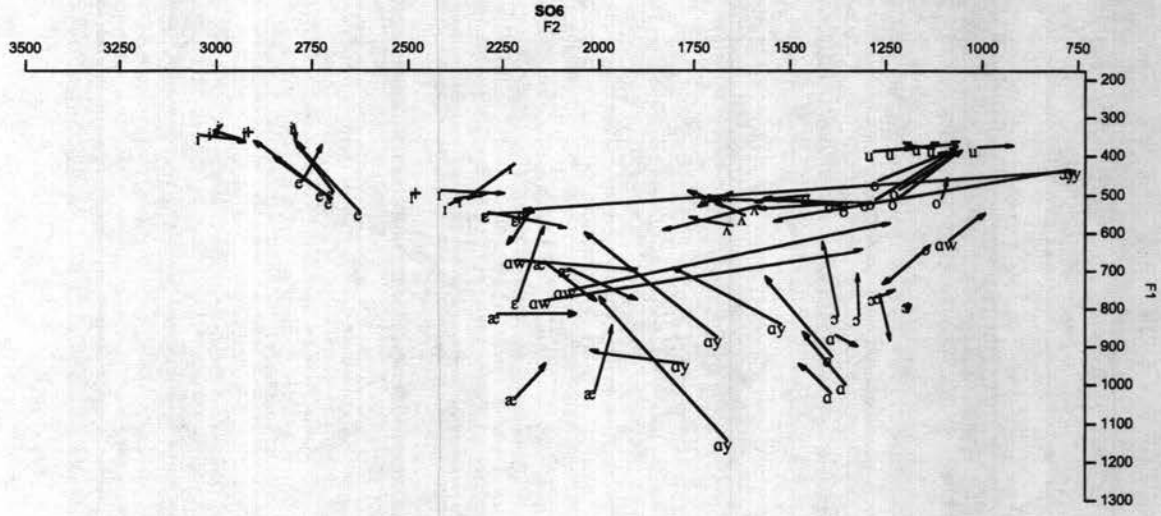

MI8
F2
3500 3250 3000 2750 2500 2250 2000 1750 1500 1250 1000 750
200 300 400 500 600 700 800 900 1000 1100 1200 1300
F1
MI9
F2
3500 3250 3000 2750 2500 2250 2000 1750 1500 1250 1000 750
200 300 400 500 600 700 800 900 1000 1100 1200 1300
F1
MI0
F2
3500 3250 3000 2750 2500 2250 2000 1750 1500 1250 1000 750
200 300 400 500 600 700 800 900 1000 1100 1200 1300
F1

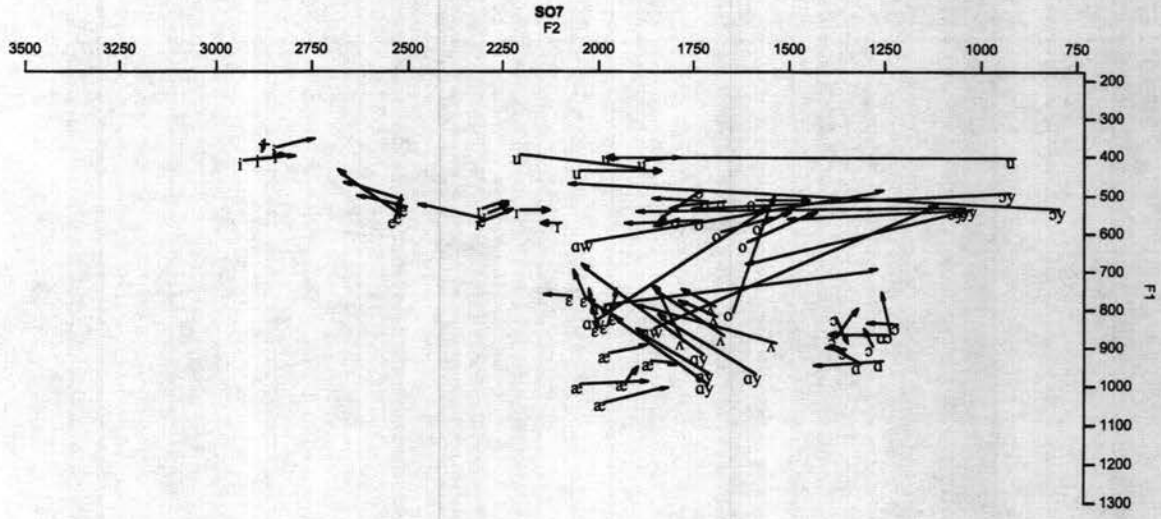
Southern Males

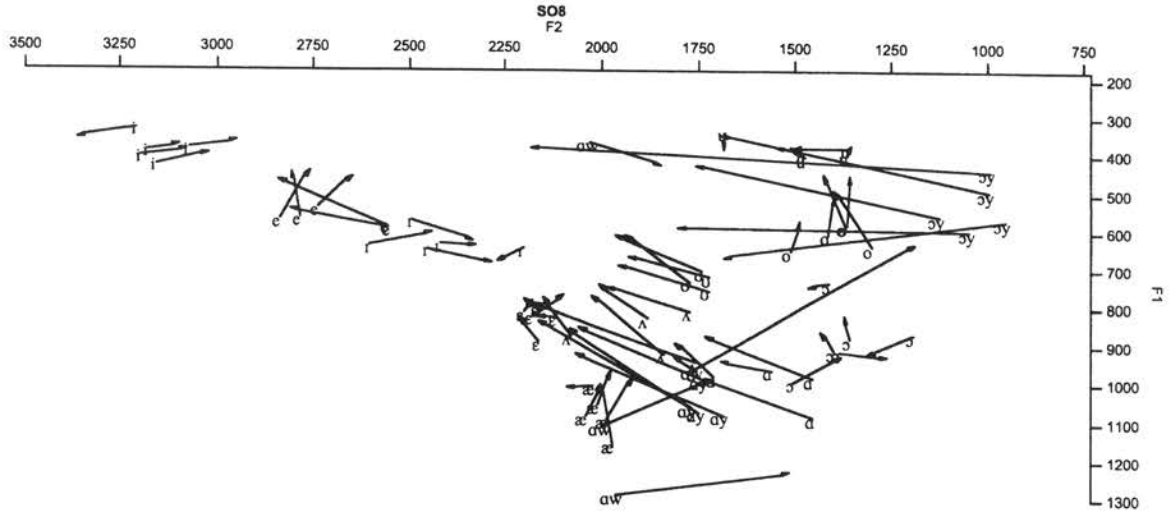


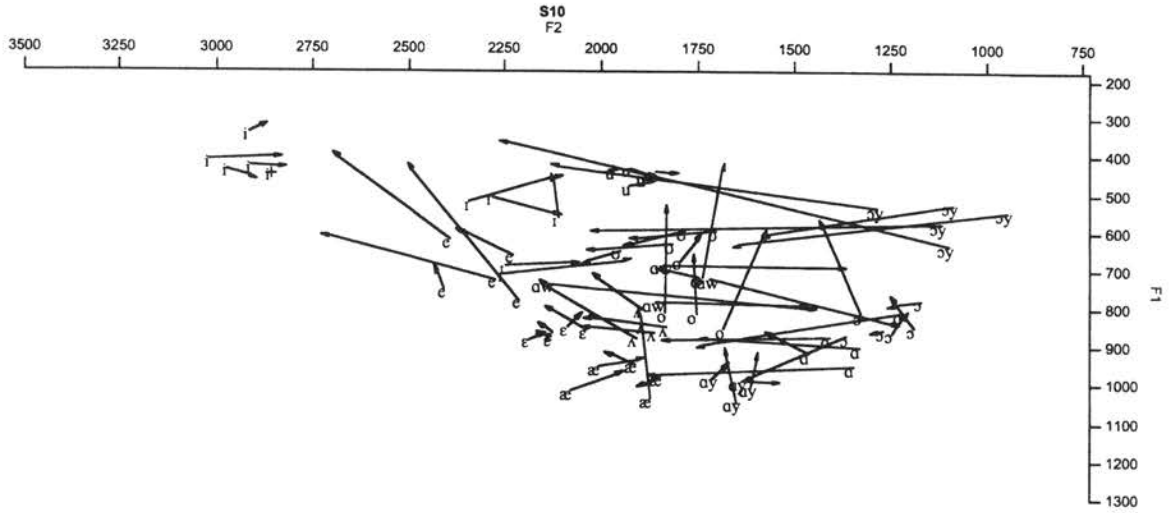


Southern Females

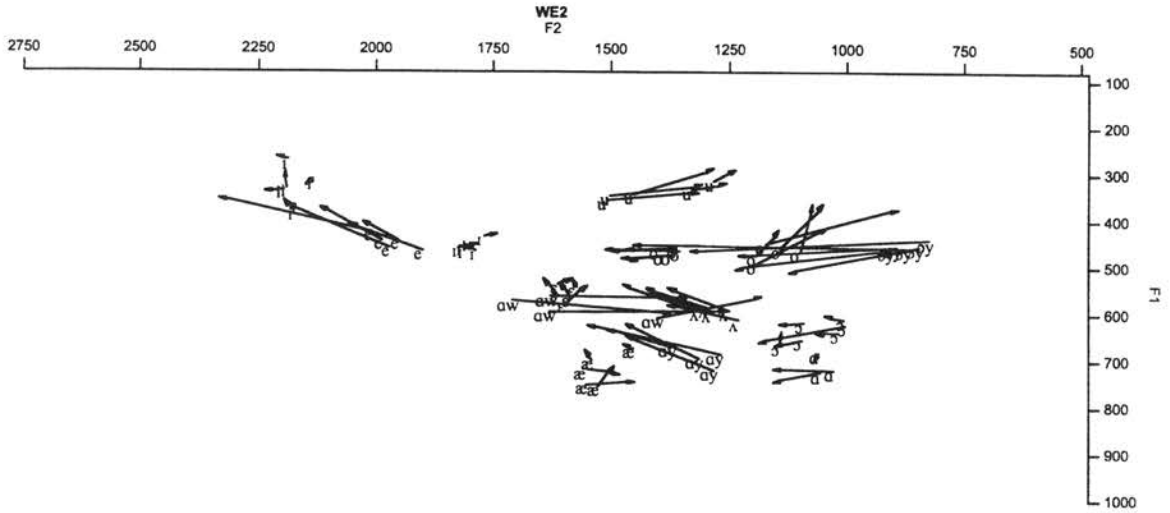


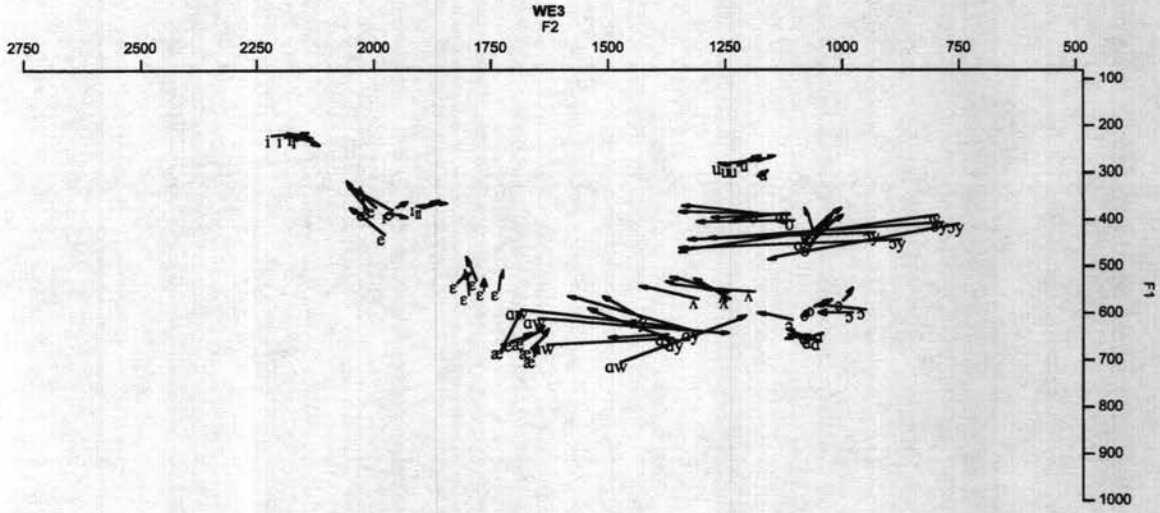


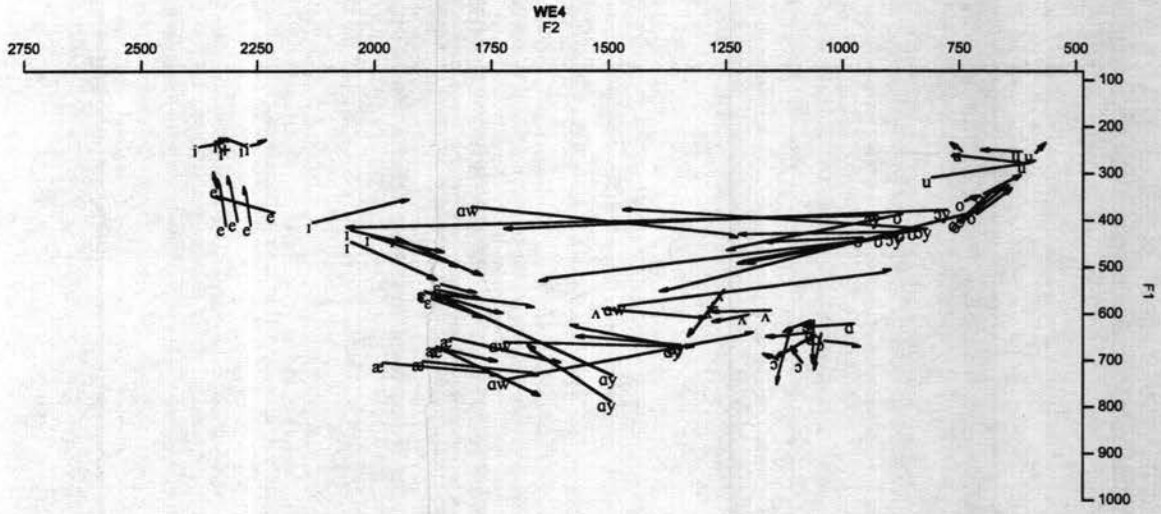


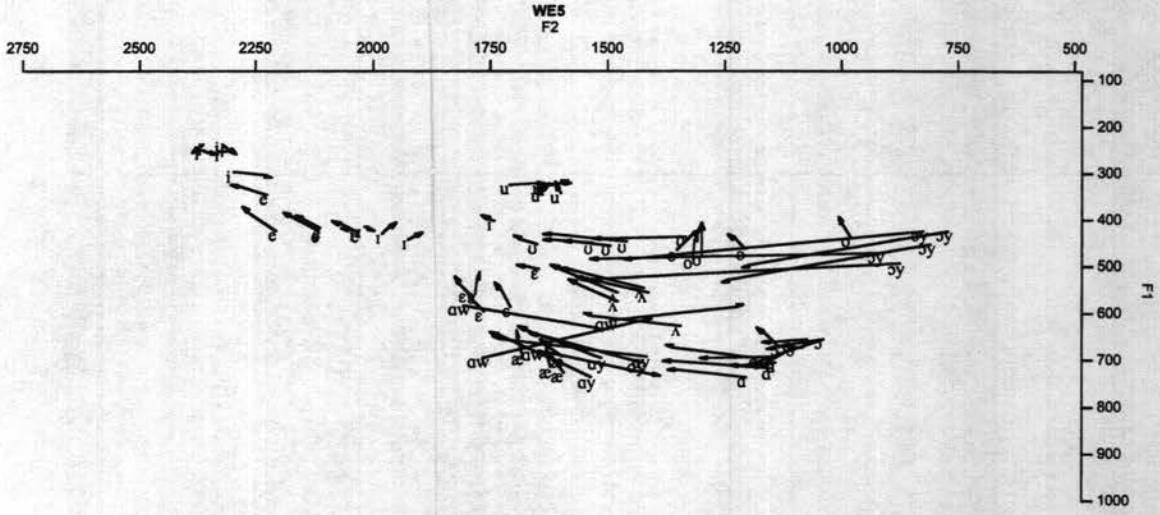


Western Males

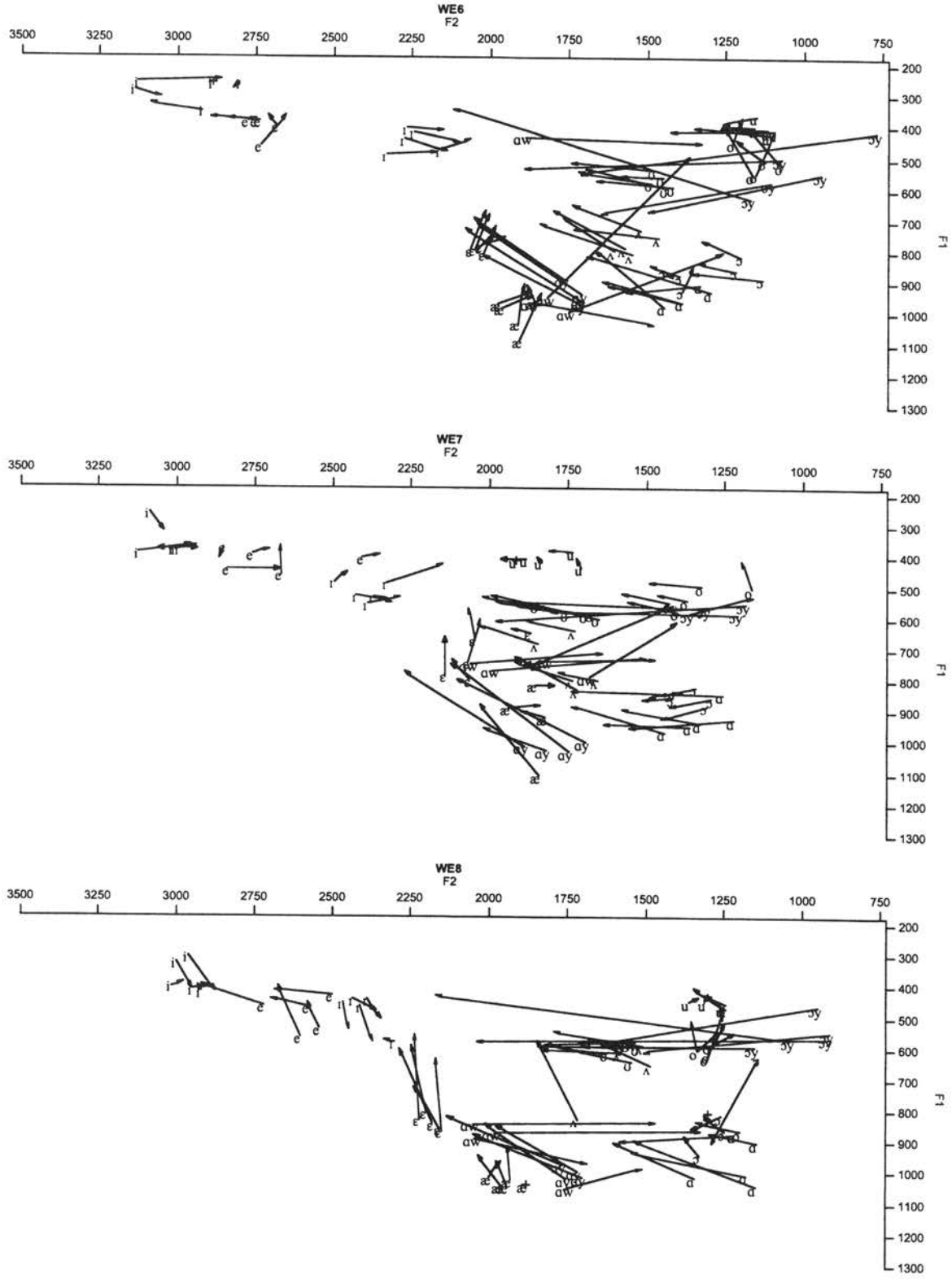


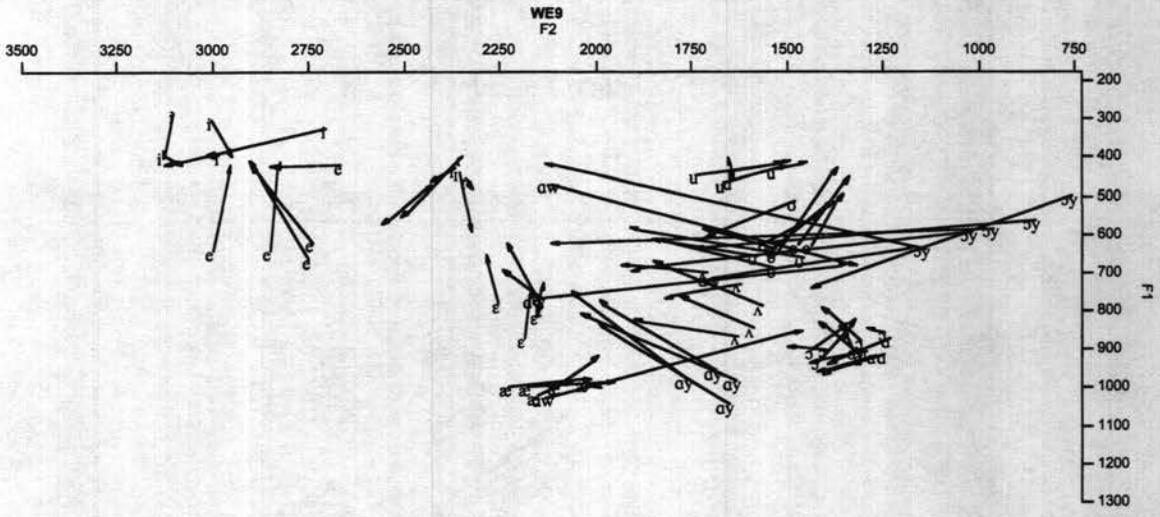

WE3
F2
2750
2500
2250
2000
1750
1500
1250
1000
750
500
100
200
300
400
500
600
700
800
900
1000
F1


WE4
F2
2750
2500
2250
2000
1750
1500
1250
1000
750
500
100
200
300
400
500
600
700
800
900
1000
F1


WE5
F2
2750
2500
2250
2000
1750
1500
1250
1000
750
500
100
200
300
400
500
600
700
800
900
1000
F1

Western Females




WE9
F2
3500
3250
3000
2750
2500
2250
2000
1750
1500
1250
1000
750
200
300
400
500
600
700
800
900
1000
1100
1200
1300
F1

APPENDIX C: STIMULUS MATERIALS FROM THE NATIONWIDE SPEECH PROJECT CORPUS

Talker	Token	Sentence	6AFC	FreeClass
A18	23	Kill the bugs with this spray.	✓	
A18	34	Please wipe your feet on the mat.	✓	✓
AT1	13	He rode off in a cloud of dust.	✓	
AT1	37	Ruth poured herself a cup of tea.	✓	✓
AT2	1036*	Ruth had a necklace of glass beads.	✓	
AT2	49	The car drove off the steep cliff.	✓	✓
AT3	26	My son has a dog for a pet.	✓	
AT3	29	Paul hit the water with a splash.	✓	✓
AT5	4	Banks keep their money in a vault.	✓	✓
AT5	1054*	The cow gave birth to a calf.	✓	
AT6	18	I ate a piece of chocolate fudge.	✓	✓
AT6	1063*	The glass had a chip on the rim.	✓	
AT7	7	Cut the meat into small chunks.	✓	
AT7	65	The heavy rains caused a flood.	✓	✓
AT9	90	Throw out all this useless junk.	✓	
AT9	97	Watermelons have lots of seeds.	✓	✓
M12	51	The chicken pecked corn with its beak.	✓	
M12	69	The lion gave an angry roar.	✓	✓
MI0	21	I've got a cold and a sore throat.	✓	
MI0	71	The nurse gave him first aid.	✓	✓
MI1	1062*	The gambler lost the bet.	✓	
MI1	74	The poor man was deeply in debt.	✓	✓
MI3	14	He was scared out of his wits.	✓	✓
MI3	57	The detectives searched for a clue.	✓	
MI4	6	Break the dry bread into crumbs.	✓	
MI4	79	The stale bread was covered with mold.	✓	✓
MI6	85	The wedding banquet was a feast.	✓	
MI6	93	To store his wood, he built a shed.	✓	✓
MI8	17	He's employed by a large firm.	✓	✓
MI8	43	The bloodhound followed the trail.	✓	
MI9	48	The candle burned with a bright flame.	✓	
MI9	82	The swimmer dove into the pool.	✓	✓
NE0	9	Follow this road around the bend.	✓	✓
NE0	99	We heard the ticking of the clock.	✓	
NE1	2	A spoiled child is a brat.	✓	
NE1	77	The shepherds guarded their flock.	✓	✓
NE2	1066*	The house was robbed by a thief.	✓	✓
NE2	73	The pond was full of croaking frogs.	✓	
NE3	50	The car was parked at the curb.	✓	✓
NE3	59	The firemen heard her frightened scream.	✓	
NE4	42	The bird of peace is the dove.	✓	
NE4	94	Tree trunks are covered with bark.	✓	✓

NE6	27	Old metal cans were made with tin.	✓	✓
NE6	101	Your knees and your elbows are joints.	✓	
NE7	58	The doctor prescribed the drug.	✓	✓
NE7	92	To open the jar, twist the lid.	✓	
NE8	22	Keep your broken arm in a sling.	✓	
NE8	95	Unlock the door and turn the knob.	✓	✓
NO0	1041*	That job was an easy task.	✓	
NO0	91	Tighten the belt by a notch.	✓	✓
NO2	47	The cabin was made of logs.	✓	
NO2	88	They tracked the lion to his den.	✓	✓
NO3	32	Peter dropped in for a brief chat.	✓	✓
NO3	96	Wash the floor with a mop.	✓	
NO4	35	Raise the flag up the pole.	✓	✓
NO4	55	The cut on his knee formed a scab.	✓	
NO5	76	The shepherd watched his flock of sheep.	✓	✓
NO5	80	The story had a clever plot.	✓	
NO6	30	Paul took a bath in the tub.	✓	
NO6	83	The swimmer's leg got a bad cramp.	✓	✓
NO8	60	The flashlight casts a bright beam.	✓	✓
NO8	70	The mouse was caught in the trap.	✓	
NO9	5	Bob was cut by the jacknife's blade.	✓	✓
NO9	31	Paul was arrested by the cops.	✓	
S10	64	The guests were welcomed by the host.	✓	
S10	100	We swam at the beach at high tide.	✓	✓
S22	33	Playing checkers can be fun.	✓	✓
S22	40	Spread some butter on your bread.	✓	
SO1	0	A bicycle has two wheels.	✓	
SO1	68	The landlord raised the rent.	✓	✓
SO2	25	My jaw aches when I chew gum.	✓	✓
SO2	1078*	The sick child swallowed the pill.	✓	
SO5	11	Get the bread and cut me a slice.	✓	✓
SO5	75	The scarf was made of shiny silk.	✓	
SO6	67	The judge is sitting on the bench.	✓	✓
SO6	98	We camped out in our tent.	✓	
SO7	12	Greet the heroes with loud cheers.	✓	
SO7	45	The bride wore a white gown.	✓	✓
SO8	10	For dessert he had apple pie.	✓	✓
SO8	38	She cooked him a hearty meal.	✓	
WE2	28	Our seats were in the second row.	✓	
WE2	39	She shortened the hem of her skirt.	✓	✓
WE3	52	The chicks followed the mother hen.	✓	
WE3	72	The plow was pulled by an ox.	✓	✓
WE4	15	Her entry should win first prize.	✓	
WE4	56	The dealer shuffled the cards.	✓	✓
WE5	1001*	A round hole won't take a square peg.	✓	✓
WE5	19	Instead of a fence, plant a hedge.	✓	
WE6	44	The boat sailed across the bay.	✓	✓

WE6	46	The burglar escaped with the loot.	✓	
WE7	3	Ann works in the bank as a clerk.	✓	
WE7	16	Her hair was tied with a blue bow.	✓	✓
WE8	81	The super highway has six lanes.	✓	✓
WE8	86	The witness took a solemn oath.	✓	
WE9	20	It was stuck together with glue.	✓	✓
WE9	84	The thread was wound on a spool.	✓	

*Token numbers higher than 1000 indicate that the stimulus item was repeated by the talker and that the second repetition was used in the present study.

APPENDIX D: STIMULUS MATERIALS FROM THE TIMIT CORPUS

New England

Talker	Token	Sentence
cpm0	384	Please shorten this skirt for Joyce.
dab0	139	The bungalow was pleasantly situated near the shore.
dac1	451	The thick elm forest was nearly overwhelmed by Dutch elm disease.
jeb1	837	Computers are being used to keep branch inventories at more workable levels.
pgh0	924	They consider it simply a sign of our times.
pgr0	420	Seamstresses attach zippers with a thimble, needle, and thread.
psw0	257	We are open every Monday evening.
stk0	34	Don't do Charlie's dirty dishes.
tjs0	1822	But to the infuriation of scientists, for no known reason, not all of them did.
tpf0	605	We would lose our export markets and deny ourselves the imports we need.
trr0	918	We know that actors can learn to portray a wide variety of character roles.

North

Talker	Token	Sentence
dbp0	168	Who took the kayak down the bayou?
jar0	728	Poach the apples in this syrup for twelve minutes, drain them, and cool.
jpm0	1998	She took it grudgingly, her dark eyes baleful as they met his.
jrp0	585	Our entire economy will have a terrific uplift.
pgl0	469	In fact, our whole defensive unit did a good job.
ppc0	152	Mosquitoes exist in warm, humid climates.
rcw0	1371	She used these new ways in daily life as the last step.
rjm0	1858	She drank greedily and murmured, "Thank you," as he lowered her head.
rlr0	566	Contrast trim provides other touches of color.
rms0	1113	Whether historically a fact or not, the legend has a certain symbolic value.
wew0	101	Kindergarten children decorate their classrooms for all holidays.

North Midland

Talker	Token	Sentence
cef0	1765	No other visitor inquired for her that evening.
dwm0	106	A huge tapestry hung in her hallway.
hmr0	1119	The record teems with romance and adventure.
jjb0	419	The football team coach has a watch thin as a dime.
mjb1	418	He murmured to himself with firmness, "No surrender."
msm0	476	So if anybody solicits by phone, make sure you mail the dough to the above.
ree0	294	Should giraffes be kept in small zoos?
rjb1	300	Amoebas change shape constantly.
rwa0	253	By eating yogurt, you may live longer.

tlb0	234	The paperboy bought two apples and three ices.
tpg0	393	She uses both names interchangeably.

South Midland

Talker	Token	Sentence
css0	750	Add a few caraway seeds, too, if you'd like.
dls0	1628	"This is my hand-ledger," he informed him in an absorbed way.
esg0	252	The gorgeous butterfly ate a lot of nectar.
gag0	645	He ignores guidebook facts.
jee0	67	Last year's gas shortage caused steep price increases.
jmm0	445	What is this large thing by the ironing board?
jrh0	1755	Only then did he decide he didn't want one.
lel10	1876	A voice spoke near at hand.
pcs0	189	Destroy every file related to my audits.
trc0	589	One of the problems associated with the expressway stems from the basic idea.
trt0	57	The prowler wore a ski mask for disguise.

South

Talker	Token	Sentence
chl0	267	Draw every outer line first, then fill in the interior.
ctt0	2188	He says he'll be here on the one o'clock plane.
gsh0	6	Bright sunshine shimmers on the ocean.
hmg0	105	You must explicitly delete files.
jpg0	291	The avalanche triggered a minor earthquake.
jwg0	2155	He was kneeling to tie his shoelaces.
ram0	375	Growing well-kept gardens is very time consuming.
rew1	870	It will accommodate firing rates as low as half a gallon an hour.
sas0	296	The two artists exchanged autographs.
srr0	501	But this doesn't detract from its merit as an interesting, if not great, film.
wch0	362	The annoying raccoon slipped into Phil's garden every night.

West

Talker	Token	Sentence
bar0	689	Men believed they could control nature by obeying a moral code.
bbr0	1685	Come on, let's hurry down before they lock up for the day.
bml0	359	The system may break down soon, so save your files frequently.
cth0	579	Usually, they titter loudly after they have passed by.
dlf0	143	The sound of Jennifer's bugle scared the antelope.
dlr0	1929	Oilfield workers were a rough, tough lot.
dlr1	333	An adult male baboon's teeth are not suitable for eating shellfish.
hbs0	405	Of course you can have another tuna fish sandwich.
jai0	164	Curiosity and mediocrity seldom co-exist.
klr0	339	Did Sean catch that big goose without help?
ntw0	348	I'll have a scoop of that exotic purple and turquoise sherbet.