Social and Linguistic Conditioning of the Sociolinguistic Variable (ai) among Textile Mill Workers of Columbus, Georgia, and Southeast Alabama

by

Elizabeth Ann Topping

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Arts

Auburn, Alabama
May 14, 2010

Keywords: Sociolinguistic Variable (ai), Monophthong, Southern American English

Approved by

Robin Sabino, Chair , Associate Professor of English
Thomas Nunnally, Associate Professor of English
Bert Hitchcock, Hargis Professor of English

Abstract

The purpose of this study is to examine the the social and linguistic conditioning as well as the frequency of the sociolinguistic variable (ai) among a group of European-American textile mill workers born in Columbus, Georgia, and Southeastern Alabama between 1896 and 1935. Fifteen interviews are transcribed with each possible instance of (ai) coded for the relevant external factor groups of birthplace, birthdate, and gender as well as the internal factor groups of following phonological environment, word frequency, word class, and word stress.

Multivariate, bivariate, and univariate analyses conducted using JMP 8.0.1 identify the significantly conditioning factor groups on production of the monophthong in this corpus. Six analyses are conducted. In the final analysis, only birthplace, gender, and following phonological environment significantly condition production of the monophthong. Tokens of *I* and containing the roots *like* and *right* are excluded and analyzed separately because they affect results.

Overall, the speakers included in this study produce the monophthong at a very high rate. The analysis conducted after exclusion of the tokens of *I* lowers this rate from 92.60 percent to 89.38 percent, but further analysis after excluding tokens containing the roots *like* and *right* confirms the initial finding. The final logistic regression indicates that the speakers in this study produce the monophthong at a rate of 94.71 percent.

Acknowledgments

I would first like to thank my major advisor, Dr. Robin Sabino, for her constant and unwavering support, patience, knowledge, and encouragement throughout this process. This project could not have been completed without her. I would also like to thank the members of my committee for their time, contributions, and encouragement: Dr. Thomas Nunnally and Dr. Bert Hitchcock. I would also to thank the Columbus State University Archives for allowing me access to the interviews that form the corpus of this study. I also want to thank my family for their encouragement and support. And to Richard, who has constantly encouraged and motivated me.

Table of Contents

List of Tables

List of Figures

I. Introduction

The definition and development of Southern speech have been of interest to linguists for decades. Of the more than 3,800 entries in McMillan and Montgomery's 1989 annotated bibliography of Southern American English research, over 600 focus on Southern phonology. Many of these studies focus on one of the most salient phonological features of Southern American English, the focus of this study, the pronunciation of (ai) as a monophthong rather than a diphthong, or *time* as [ta:m].

One area of considerable interest is the origin of Southern English. One theory offered for the development of Southern speech, arising out of traditional dialectology studies, suggests that the major varieties of American English, including Southern English, developed during the colonial period, directly reflecting settlement areas (Kurath, 1949, cited in Bailey, 1997). Researchers in this camp also suggest that the presence of large numbers of African Americans living near and with European Americans during the colonial period more than likely influenced the development of Southern American English (Feagin, 1997; Mufwene, 2003). Schneider (2003) also states that African American English is closely related to Southern American English.

A very different position holds that, instead, Southern English as now known did not develop until the 1870s, when the South emerged from the Reconstruction period. Bailey and Tillery (1996) suggest that many features associated with modern Southern

American English, including monophthongal (ai), were not widespread in Texas prior to 1875.

Though traditional dialectology premises suggest that this hallmark Southern feature has long been present, Bailey and Tillery (1996) argue that evidence from Texas speech shows that, instead, modern Southern American English developed after the Civil War. Tillery and Bailey (2003) point to accelerating urbanization beginning about 1880 in the historically rural South. Indeed, Labov (1990, p. 206) points out that "sudden changes in linguistic systems are associated with catastrophic social events." However, Texas was settled later than were many parts of the American South. When the area opened for settlement in the 1830s and 1840s, many migrants came through and from Alabama (Otto, 1989). Thus, research on the speech of people born in areas farther eastward than Texas should be investigated to test Bailey and Tillery's proposal.

In investigating the factor groups conditioning the alternation between these variants, sociolinguistic researchers have studied the social factor groups of speakers' socioeconomic status, ethnicity, age, gender, and origin. Researchers have also found various linguistic factor groups such as word stress, word frequency, word class, and following phonological environment, to affect production of the variants. As part of the ongoing research on Southern English, this study investigates the frequency and sociolinguistic conditioning of the sociolinguistic variable (ai) among European-American textile mill workers from southeastern Alabama and Columbus, Georgia, born between 1896 and 1935.

In 1828, the Georgia state legislature voted to establish a city at the fall line along the eastern bank of the Chattahoochee River on the Georgia-Alabama state line. The town was laid out physically that year, with cotton-trading facilities, gristmills, and sawmills appearing shortly thereafter. The Creek Indians remained on the west side of the Chattahoochee River until 1836, at which time they were forcibly removed ("Creek Indians," 2002; Green, 1982).

Two years later, the first textile mill opened north of town. During the Civil War, Columbus became one of the top five manufacturing centers of the Confederacy, with textile mills and other heavy industry. Although all these facilities were burned shortly after the official end of the Civil War, rebuilding began almost immediately. Textiles mills were operating by December of 1865, less than a year after their destruction.

The 1873 economic crash and ensuing recession caused hardships to non-textile industries, but the textile industry flourished, becoming the area's major employer. These mills expanded production through the 1870s, catapulting the city to one of the nation's top producers of textiles by 1880 ("Columbus," 2004).

In 1900, the Bibb Manufacturing Company purchased land north of Columbus to build a mill and mill housing. This company eventually became one of Georgia's largest employers. The mill town incorporated in 1909 as Bibb City. Over time, Columbus grew around the mill village, but the town retained its separate political identity until 2000, when the Bibb mill closed and the town merged with Columbus. ("Bibb Manufacturing Company," 2006).

Employees of the textile industry witnessed and experienced many of the tumultuous domestic events of the twentieth century. The rise of the textile industry in Columbus mirrored the rise of the textile industry across the South as an important segment of the economy, making Columbus ideal for testing Bailey and Tillery's (1996) hypothesis concerning the influence of increasing urbanization on the monophthongal variant [a:]. Increasing demand for textiles during World War I led to higher wages generally, but those wages fell as hours increased shortly after the end of World War I. Changing working conditions led to a series of labor strikes across the nation in the 1920s and the mid-1930s (Hall, J. D., Leloudis, J., Korstad, R., Murphy, M., Jones, L., & Daly, C. B., 1987). During World War II, many of the Columbus textile mills contracted with the federal government to supply soldiers' needs. Although some employees were granted deferments of their required military service, mills were shorthanded. Management looked outside the local community in order to keep their machines running, recruiting heavily from the Houston County area of southeastern Alabama ("Columbus," 2004; "The Archives: Mill Workers Oral Histories," n.d.).

When I moved to Columbus in 1986, textiles were still a central industry, and I interned at a local carpet mill while a college undergraduate in the later 1990s. Since then, most of the local textile facilities, cotton and carpet, have closed, and many of the structures no longer stand. However, residents have initiated a number of projects to preserve Columbus's history, including ongoing work to add Bibb City to the National Historic Register ("The Archives," n.d.; Barnes, 2008) and recording interviews with residents to preserve their memories of the area. One of these oral history projects, conducted specifically with former Columbus textile mill employees, provides the data for this study.

II. Literature Review

Introduction to Southeastern Alabama and Columbus

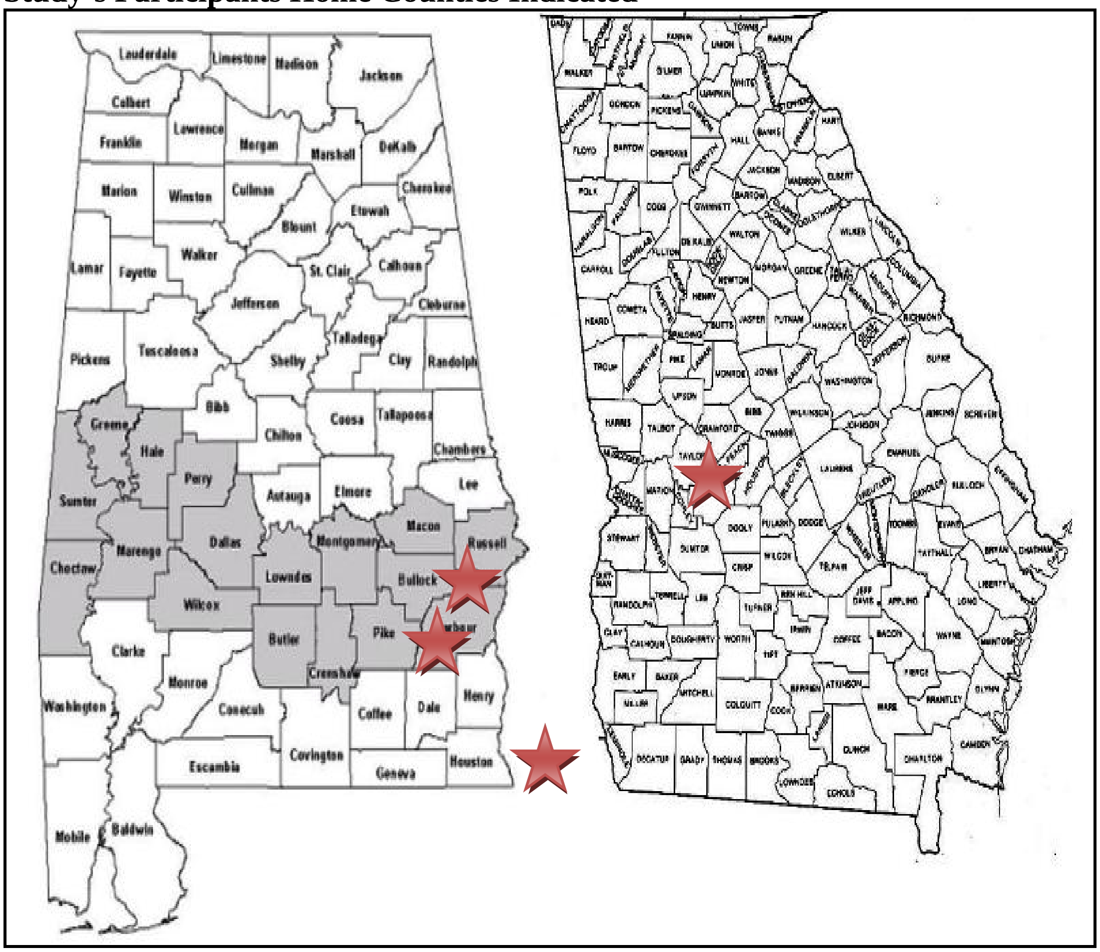
The present study focuses on the production of the sociolinguistic variable (ai) by speakers from southeastern Alabama and Columbus, Georgia. Speakers included in this study were born in one of three Alabama counties: Bullock, Pike, and Houston. Bullock is traditionally considered a Black Belt County, and Houston is a Wiregrass county. Some sources consider Pike a Black Belt county; others claim Pike as a Wiregrass county.[1] As shown in Map 1 below, Bullock County lies farthest north and Houston County farthest south, with Pike County in between them ("Archives of Wiregrass History & Culture," n.d.; BlackBelt Action Commission, 2004).

In the antebellum period, Black Belt counties, such as Bullock, relied heavily on plantation agriculture. For example, an early report described Union Springs in Bullock County as "a healthy land where lived the wealthiest plantations" ("Union Springs," 2006). The Wiregrass counties contained sandy soil considered poor and thus were home to fewer large-scale agricultural ventures such as the cotton plantations, which utilized enslaved labor ("The City of Dothan, Alabama," 2006). Though agriculture has been historically important to all three Alabama counties, the settlers and therefore economic focus of these areas are different from each other.

[1] According to the Archives of Wiregrass History and Culture (n.d.), the Wiregrass counties include Barbour, Coffee, Covington, Crenshaw, Dale, Geneva, Henry, Houston, and Pike counties.

Columbus has a different history from these counties in Alabama. Planned as a trading town due to its location at the northernmost navigable point on the Chattahoochee River, Columbus became dominated by the textile industry over time ("Columbus," 2004). Though the city is in an area historically agriculturally oriented, Columbus's economy has not had a primarily agricultural focus. Figure 1 below shows Alabama and Georgia with the home counties of the study's participants indicated.

Figure 1: Labeled County Maps of Alabama and Georgia, Black Belt Shaded, Study's Participants Home Counties Indicated



Adapted from "Traditional Counties of the Black Belt" produced by Center for Business and Economic Research (n.d.) and map of Georgia from the Selig Center for Economic Growth (n.d.)

Sociolinguistic Variable (ai): Origin

The sociolinguistic variable (ai) is considered a hallmark feature of Southern English. However, there is debate as to the origin of this feature. Traditional dialectology suggests that Southern English developed during the colonial period, directly reflecting settlement patterns (Kurath, 1949, cited in Bailey, 1997), thus implying that the monophthongal (ai) has long been present in Southern English. However, other researchers believe that instead modern Southern English did not develop until the 1870s. Bailey and Tillery (1996) investigate a number of phonological and grammatical features typical of Southern English, suggesting that many features, including the monophthong [a:] were not present in Texas prior to 1875. In particular, their analysis of the speech of three Texans, born in 1847, 1897, and 1976, documents a change from a fully diphthongal [ai] to a fully monophthongal [a:]. Bailey and Tillery argue that their results show that modern Southern American English developed in the last quarter of the nineteenth century.

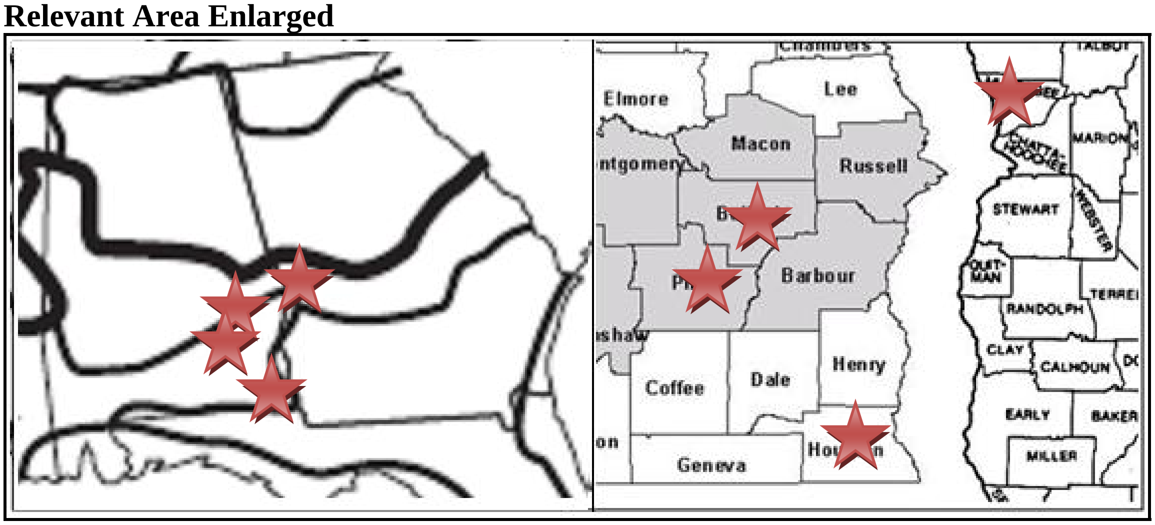
Researchers who focus on the social and linguistic conditioning of the sociolinguistic variable (ai) consistently find the social factor groups of socioeconomic status, ethnicity, age, gender, and region to correlate with the production of the monophthongal variant. The linguistic factor groups reported in the literature that correlate with production of the monophthong include following phonological environment, word frequency, word class, and word or syllable stress.

External Variables

Although the monophthong is typical in Southern and Southern-derived speech varieties, within the American South, (ai) patterns differently in different locales.

Dialectology studies delineate a number of dialect divisions within the American South based on phonological, morphological, and lexical items. In their LAGS research, Pederson, McDaniel, Bailey, and Bassett (1986) divide the Gulf States into two primary areas, the Interior and the Coastal. Speakers included in the present study were born in the Coastal region, which is divided further. The Coastal Black Belt, a transitional area, stretches from the Atlantic coast of Georgia across central Alabama. Columbus, Bullock County, and northern Pike County lie within the Coastal Black Belt. Farther south, the Central Piney Woods encompasses Houston County and southern Pike County (see Figure 1). Figure 2 below shows the primary LAGS divisions and an enlarged view of Figure 1. The Coastal Black Belt is labeled as section A, and the Central Piney Woods as section B, to the south.

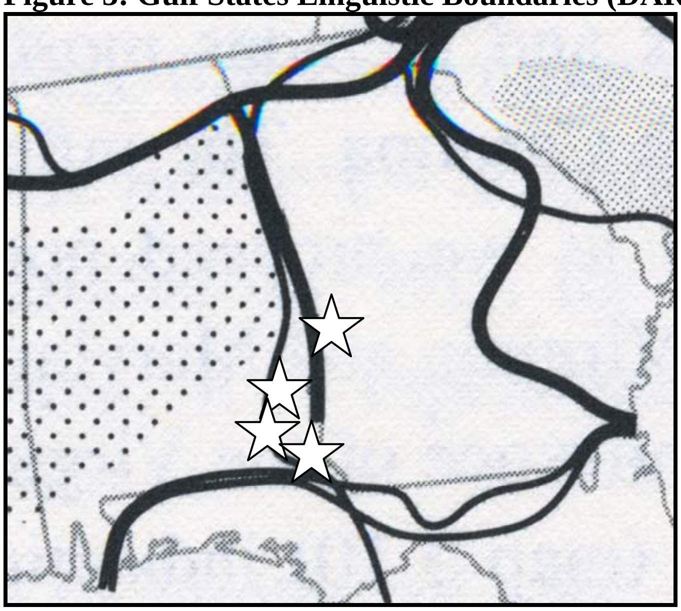
Figure 2: LAGS Dialect Divisions and County Map of Alabama and Georgia, Relevant Area Enlarged



Adapted from Lance (1994) with approximate county locations starred, Center for Business and Economic Research (n.d.), and Selig Center for Economic Growth (n.d.)

Analyzing primarily lexical data in his DARE study, Carver (1987) establishes the Chattahoochee River, immediately west of Columbus, as a linguistic boundary between Georgia and Alabama. Within Alabama, he divides the state into three regions: the Alabama Appalachian area; the southern and southeastern area, including the Wiregrass area; and the central area, including the Black Belt. The central area was settled much later than the other two areas due to the continuing presence of Native American tribes until the late 1830s. Figure 3 below shows the DARE linguistic boundaries, with the central Alabama region dotted and the Chattahoochee River boundary shown in a thick bolded line. Approximate locations of the four areas discussed are starred.

Figure 3: Gulf States Linguistic Boundaries (DARE)



Adapted from Carver (1987)

Schneider (1998) also analyzes LAGS lexical data from western Georgia and eastern Alabama. He suggests that there are some differences between the two states, but that there is little support for the Chattahoochee as a linguistic boundary. Instead, he sees the boundary as "a broad zone of gradually changing lexical choices" (p. 144).

In addition to the divisions suggested by dialectology studies, quantitative sociolinguists report that origin and location influence speakers' choice between the diphthong and monophthong. McNair (2005), in her study of mill villagers and farmers living in Griffin, Georgia, reports that the mill villagers, whose families migrated from Appalachia, produce the monophthong more frequently than do the farmers, whose families migrated from coastal areas of the Southeast. Thomas (1997) reports a contrast developing between metropolitan and rural areas in Texas, with metropolitan-area speakers more likely to produce the diphthong.

Origin is not the only social variable influencing the alternation between the monophthong and the diphthong. Socioeconomic status is reported as conditioning the variation. Crane (1977) reports that in Tuscaloosa, Alabama, the members of the highest socioeconomic class are least likely to use the monophthongal variant, and Edwards (1997) reports that working-class African Americans in Detroit who retain a large number of Southern features are likely to use the monophthongal variant. However, Head's (2003) study of forty-two speakers in Elba, Alabama, a community close to Houston County, reports the distinction between middle and working class not to be significant.

Ethnicity also plays a role in production of the monophthongal (ai), and like findings for socioeconomic status, results are community-specific. Feagin (2003) observes that African Americans generally avoid sound changes observed in European-

American communities, which suggests that production of the monophthong may differ between the two ethnicities. Indeed, Head (2003) finds that African Americans are more monophthongal than European Americans in Elba, Alabama, though other elements have a greater influence on the alternation. Fridland (2003) reports that African Americans in Memphis, Tennessee, lead the weakening of the diphthong to a monophthong in all phonetic contexts. In contrast, other studies find that African Americans are significantly less likely than European Americans to produce a monophthong before voiceless obstruents (Bailey & Thomas, 1994; Schilling-Estes, 2000; Thomas & Bailey, 1998). In a community containing the three ethnic groups European Americans, African Americans, and bilingual Cherokee, Anderson (1999) reports that each group demonstrates different patterns in production of the monophthong.

Age is another variable consistently found to correlate with production of the monophthongal variant. Unlike socioeconomic status and ethnicity, whose effects differ from community to community, older speakers typically produce the monophthong more frequently than younger speakers do. Fridland (2003) suggests that this choice reflects younger speakers trying to distance themselves from what they perceive to be strongly Southern speech. Head (2003) reports that older speakers produce the monophthong more frequently before voiceless obstruents, which may also suggest distancing by younger speakers. Bailey and Bernstein (1989), Bowie (2001), Crane (1977), Edwards (1997), Fridland (2003), and McNair (2005) report this age-related effect. However, some research suggests that younger speakers are leading the increased production of the monophthong before voiceless obstruents. This difference between older and younger speakers has been found among speakers in Southern Appalachia (Thomas, 2001) as well as among African American residents of Detroit who have historically maintained close ties with European-American migrants who moved to the city from the Appalachian region (Anderson, 2002).

Gender is also an important conditioning factor in sociolinguistic variation. Labov (1990) reminds us that in instances of stable sociolinguistic variation, males use nonstandard forms more frequently than females do. For example, research into the stable alternation between [n] and [ŋ] in words ending with *-ing* has revealed that males are significantly more likely to choose [n] and females [ŋ]. In instances in which an incoming variant holds greater prestige than do the local variants, or change from above, females tend to favor the incoming prestigious variant. Labov (1966, cited in Chambers, 2002) documents this change in the adoption of the pronunciation of (r) in New York City. In a study specifically related to the variable investigated in this study, Bowie (2001) reports that females are leading the shift from the monophthong to the diphthong, an incoming prestigious variant, in a community in Maryland. Finally, females are also typically the innovators in instances of change from below, when the variants do not have overt social prestige or stigma. Labov (1990) reports several examples, including research by Bailey and Bernstein (1989) using the Texas Poll data of 1989 in which females lead males in several cases, such as the merger of /ɔ/ and /a/ in words like *cot* and *caught*.

With respect to the sociolinguistic variable (ai), some researchers report that males produce the monophthong more frequently than females do. Edwards (1997) reports that working-class African American males produce the monophthong more frequently than females do in Detroit, Michigan. Fridland (2003) reports a similar finding for Memphis, Tennessee, though males are "showing a slight lead" (p. 289)[2]. These results suggest that the alternation between [ai] and [a:] is a case of stable variation. Interestingly, Head (2003) reports no significant difference between males and females in her analysis.

Internal Variables

In addition to the social variables discussed above, researchers have identified a number of linguistic variables that affect the production of the monophthong. Specifically, following phonological environment, word frequency, word class, and stress condition the alternation between the monophthong and the diphthong; following phonological environment has received the greatest attention in the literature.

Generally, following voiced obstruents are reported to favor production of the monophthong, while voiceless obstruents disfavor its production (Anderson, 1999, 2002; Bowie, 2001; Eckert, 1996; Fridland, 2003; Hazen, 2000; Labov & Ash, 1997). As shown in Table 1, Bowie (2001), Eckert (1996), and Hazen (2000) report similar conditioning effects for following phonological environment, with liquids favoring the monophthong the most and voiced obstruents the least. In contrast, Head (2003), who reports a high percentage of the monophthong, finds less differentiation between following phonological environments, though she does report that voiceless obstruents are the least favoring environment.

---

[2] Fridland (2003b) reports the raw percentages of the monophthong among both African-American and European-American speakers, which differ by only a few percentage points. She does not comment regarding the statistical significance of her finding.

Table 1: Summary of Previously Reported Implicational Hierarchies

Author	Implicational Hierarchy
Bowie (2001)	liquids > nasals, voiced obstruents > voiceless obstruents
Eckert (1996)	liquids > nasals, voiced obstruents > voiceless obstruents
Hazen (2000)	liquids > nasals > voiced obstruents > voiceless obstruents
Head (2003)	word boundary > voiced obstruents, liquids, nasals > glides, vowels > voiceless obstruent

Although not often examined in the research on (ai), word frequency has been found consistently to affect the diffusion of sound changes through the lexicon (Bybee, 2000, 2007; Phillips, 1998, 2000). In cases of synchronic variation as in vowel deletion (Bybee, 2000) or word final [t]/[d] deletion (Gregory, Raymond, Bell, Foster-Lussier, & Jurafsky, 1999, cited in Bybee, 2007), the most frequent words are typically affected first. However, Phillips (1998, 2000) also points out that speakers attempting to access unfamiliar, low-frequency words will default to national norms, and thus a sound change also can affect low-frequency words first. Phillips (2000) also points out that frequency effects are greatest early in the process of sound changes.

Consistent with research by Bybee (2000, 2007) and Phillips (1998, 2000), Mendoza-Denton, Hay, and Jannedy (2003) report in an analysis of (ai) in Oprah Winfrey's speech that words more highly frequent in their study's corpus are more likely to be produced as a monophthong. Head's (2003) study reports high-frequency words are less likely to be produced as a monophthong than low-frequency words are. However, her analysis eliminates both categorical and near-categorical types, all high-frequency words.

Neu (1980) reports that inclusion of the high-frequency, nearly categorical *and* significantly changed confidence intervals of her analysis; therefore, she excludes *and* in order to investigate better the patterning of final-cluster deletion. However, inclusion of *just*, also high-frequency and deleted at a much higher rate than the other tokens, does not significantly affect confidence intervals and is not excluded, suggesting that high-frequency, near-categorical forms should not be excluded from data analysis before examining their effects on the data.

In regard to word class, both Bowie (2001) and Head (2003) report that content words are more likely to be produced with the monophthong than function words are, with the exception of pronouns. However, in Head's (2003) study, all pronouns except *mine* are excluded because of their nearly categorical nature. Bowie (2001) limits the number of the pronouns *I* and *my* to ten percent of each interviewee's tokens due to their large number. Eckert (1996) eliminates tokens of *I* from her sample. Other researchers do not discuss their treatment of this category of tokens.

With respect to stress, Bowie (2001) reports that secondary stress within a word slightly favors the monophthong. In contrast, Eckert (1996) and Head (2003) do not find stress to influence significantly the production of the monophthong.

III. Methodology

Study Background

The textile industry has been an integral part of the Columbus economy and its social structure. However, in the latter third of the twentieth century, the textile industry experienced a significant decline. In Columbus, this decline eventually resulted in the closing of most of the area textile mills ("Columbus," 2004).

In an attempt to preserve the memories of those closely involved with this important part of Columbus history, a class at Columbus State University, then Columbus College, interviewed a number of textile employees of Columbus-area mills as part of an oral history project in 1988 ("The Archives," n.d.). Those interviewed granted the Columbus State University Archives permission to retain the interviews as well as to allow them to be used for scholarly purposes deemed appropriate by the Archives. The head of the Archives, Reagan Grimsley, granted permission to this author to analyze a subset of the original interviews for this study.

Human Subjects Review Board

As required for all research involving human research subjects, including retrospective studies such as this one, a protocol was submitted to the Auburn University Human Subjects Institutional Review Board (IRB). The project was approved as "Exempt."

Research Participants

Since not all participants in the original oral history project identified their age, birthplace, or ethnicity during the interviews, only those interviews containing that information are considered for the present study. Of those interviews that do contain age, birthplace, and ethnicity information, only a small number of those participants were African Americans. Because the number of available African-American interviews is too small to provide representative data, the study is limited to European-American speech. All fifteen of the European-American interviewees who identified their age, birthplace, and ethnicity were hourly employees or hourly employees who had been promoted to low-level supervisory positions within the mills, so socioeconomic status is constant. As Table 2 shows, one female (CF1) and three males (CM2, CM3, CM4) are from Columbus, two females (BF1, BF2) are from Bullock County, two males are from central Pike County (PM1, PM2), and one female (HF1) and six males (HM2, HM3, HM4, HM5, HM6, HM7) are from Houston County. All participants were born between 1896 and 1935. In addition to demographic information, Table 2 displays the total number of monophthongs produced by each participant, the total number of tokens, and the percentage of monophthongs produced by each participant.

Table 2: Participant Data

Participant	Birthplace	Gender	Birthdate	Total number of monophthongs/total number of tokens	Percentage of monophthong production
CF1	Columbus	Female	1927	234/273	85.71
CM1	Columbus	Male	1898	139/147	94.56
CM2	Columbus	Male	1912	131/151	86.75
CM3	Columbus	Male	1928	88/95	92.63
BF1	Bullock County	Female	1916	435/483	90.06
BF2	Bullock County	Female	1935	245/280	87.5
PM1	Pike County	Male	1912	492/525	93.71
PM2	Pike County	Male	1927	514/532	96.61
HF1	Houston County	Female	1906	60/62	96.77
HM1	Houston County	Male	1896	312/317	98.42
HM2	Houston County	Male	1915	474/505	93.86
HM3	Houston County	Male	1915	258/303	85.15
HM4	Houston County	Male	1915	492/526	93.54
HM5	Houston County	Male	1919	659/715	92.17
HM6	Houston County	Male	1923	251/275	91.27

Interview Topics

Interviewees discussed primarily their textile experiences, as dictated by the interests of the oral history project. Specific topics include the Great Depression, New Deal legislation that impacted the textile industry, the textile strike of 1934, World War II and the accompanying labor shortage, introduction of air conditioning into the mills, desegregation of the work force in the 1960s, ongoing modernization and job responsibilities, life in textile mill villages, female and male work roles, and other events the speakers found memorable between the 1920s and 1980s ("The Archives," n.d.).

Transcription of the Interviews

The fifteen selected interviews were transcribed using a Panasonic RR-830 cassette transcriber. Because speakers are often nervous around cassette recorders, transcription for sociolinguistic interviews usually begins partway through the interview. However, all but two of the interviews included in this study were less than an hour long, so they are transcribed and coded from the beginning of the audiotape to increase the number of tokens available for analysis. Coding of the two longer interviews began approximately one-third of the way into the audiotape in order to analyze the most natural data possible.

In the current study, the dependent variable is (ai). Each of the 5,289 tokens containing [ai] or [a:] is coded impressionistically as containing either a monophthong or a diphthong. Each token is also coded for a series of external and internal independent variables identified in the literature review. As in other quantitative sociolinguistic studies, each independent variable is identified as a factor group.

Coding of External Variables

The external factor groups considered in the present analysis are birthplace, gender, and age. Other external variables discussed in the literature review are constant for this set of interviews. The interviewees identified themselves as from Columbus, Bullock County, Pike County, or Houston County, areas that contributed to the settlement of eastern Texas. Tokens are coded according to the speakers' birthplaces. All tokens are coded as produced by a male or female. Each token also is labeled according to speaker's birth year. The external factor groups and their factors are summarized in Table 3 below.

Coding of Internal Variables

The internal factor groups are following phonological environment, word frequency, word class, and stress.

Tokens are coded according to the phonological environment following the dependent variable. Elements in this group are initially coded as a voiced obstruent, nasal, word boundary, vowel, liquid, voiceless obstruent, or glide. The factors in the word class factor group are noun, adverb, verb, adjective, pronoun, preposition, or conjunction, according to their entries in *The American Heritage Dictionary: Second College Edition* (DeVinne, 1982). For the factor group stress, each token is coded as unstressed, stressed, or hyperstressed. Stress is impressionistically assessed within its clause. Unstressed tokens are generally very short and not as loud as the other tokens. Hyperstressed tokens are the focus of their clause and thus are louder and often relatively longer than other tokens.

Each token is also coded for word frequency. Following Head (2003), this study uses frequencies assigned by *The American Heritage Word Frequency Book* (Carroll, J. B., Davies, P., & Richman, B., eds., 1971), which Sabino (2007) found to reflect frequency reported by online search engines more closely than other available frequency corpora. Because *The American Heritage Word Frequency Book* (Carroll, J. B., Davies, P., & Richman, B., eds., 1971) analyzes written text, and this study analyzes speech, there is some disparity between the frequency reported in the book and the actual frequency in this study's corpus. Speakers often discuss the years in which they worked at or retired from the different mills; their wages; and numbers, speeds, and types of machines in various departments. Following Head (2003), frequency is assigned for these words on the basis of their content morphemes (e.g. *nineteen* and *ninety)*. Additionally, because the interviews analyzed in the present study focus on the textile industry, a number textile-specific terms, such as *winders* and *dyehouse*, were recorded at a relatively high frequency. Nevertheless, they are assigned the frequencies given in *The American Heritage Word Frequency Book* (Carroll, J. B., Davies, P., & Richman, B., eds., 1971). Internal factor groups and their factors are summarized in Table 3 below.

Table 3: Factor Groups and Their Factors

Factor Group	**Factors**
Birthplace	Columbus, Bullock County, Pike County, Houston County
Gender	male, female
Birthdate	1896, 1898, 1906, 1912, 1915, 1919, 1923, 1927, 1928, 1935
Following Phonological Environment	voiced obstruent, nasal, word boundary, vowel, liquid, voiceless obstruent, glide
Frequency	as indicated by *The American Heritage Word Frequency Book*
Word Class	pronoun, noun, verb, adjective, adverb, preposition, conjunction
Stress	unstressed, stressed, hyperstressed

Hypothesis

As discussed in the literature review, the sociolinguistic literature identifies the six external variables of birthplace, socioeconomic status, ethnicity, gender, and age as correlating with production of the monophthong. However, the sample available to this study permits only the consideration of birthplace, gender, and birthdate. The literature also identifies four internal variables (i.e., following phonological environment, word frequency, word class, and stress) as correlating with production of the monophthong. The null hypothesis assumes that there are no differences among the different factors in each factor group: birthplace, gender, birthdate, following phonological environment, word frequency, word class, and stress.

Statistical Analysis

In order to evaluate the relationship between the dependent variable, (ai), and the independent variables, this study uses two statistical analysis methods common to sociolinguistic studies (Tagliomonte, 2006). The first statistical analysis method, logistic regression, evaluates the simultaneous contribution of the independent variables to the dependent variable, thus indicating how much of the observed variation is correlated with the factor groups. Chi Square evaluates the relationship between the dependent variable and the independent variables within factor groups, determining whether the differences between observed and expected frequencies of the factors are due to the influence of the independent variables or sampling errors.

Traditionally, sociolinguistic studies use Variable Rule Analysis (Cedergren & Sankoff, 1974; Rand & Sankoff, 1990) to analyze the relationship between the independent variable and the dependent variables. The present study uses a comparable commercially available statistical program, JMP Statistical Discovery from SAS, version 8, chosen both for its more conservative analysis (Sabino, Diamond, & Head, 2004) as well as its flexibility and user-friendliness.

The original data set contains 5,289 tokens of (ai) produced by 15 speakers. One hundred tokens that cannot be categorized as either a monophthong or a diphthong are excluded from analysis. Sixty-four tokens coded as hyperstressed are also excluded because hyperstress coding reflects sentential stress rather than word stress. Finally, the four words containing a following glide (i.e., *Taiwan*, *flywheels,* and two tokens of *highway*) are excluded because this group is unstable due to its small size. Exclusion of these 168 tokens leaves 5,121 for analysis. These tokens are entered into a JMP spreadsheet and coded for the dependent and independent variables.

A logistic regression is conducted of the coded data set to identify significant and non-significant factor groups. Significant factor groups with more than two factors are then analyzed using pairwise Chi Square tests to determine if factors within the group differ significantly from each other. Factors that are not significantly different from each other and share one or more attributes are recoded as a single factor. For example, if participants produce the monophthong at a statistically similar rate before voiced obstruents and nasals, both of which are stops, these two factors are recoded as voiced nasal/oral stops. After recoding, a second logistic regression is conducted with only the recoded, significant factor groups. The final logistic regression is used to test the hypothesis.

In a discussion of variation found within Texas Poll linguistic data, Bernstein (1993) states that without knowing how much internal and external factor groups contribute toward variation of dependent variables, it is impossible to know whether the independent variables adequately predict variation. JMP reports an R Square, which indicates the percentage of variation accounted for by the factor groups included in analysis. Following Bernstein's (1993) suggestion, this study reports the R Square for both the initial and final logistic regressions conducted for the data.

Alpha, the level at which results are deemed significant, is set at 0.05. Unless stated otherwise, all Chi Square probability results reported in analysis are drawn from the pairwise tests. All results are discussed in the next chapter.

IV. Results and Discussion

Analysis of All Data

Overall, the rate of production of the monophthong in the complete data set is high at 92.25 percent. Exclusion of 168 tokens leaves 5,121 tokens for analysis, which exhibit the monophthong at a 92.60 percentage rate.

Head's (2003) study excludes high-frequency and nearly categorical items from analysis. Neu (1980) suggests excluding these tokens if they skew results. Following Neu (1980) and Head (2003), two high-frequency, nearly categorical items (*time* and *by*) are identified for possible exclusion by inspecting the data set. Neu (1980) only excludes tokens that significantly affect her data. Following her analysis, the rate of production of the monophthong in data excluding tokens of *time* and *by* is compared to the rate of production of the monophthong of all data with pairwise testing. The rates of production between the two data set are not significantly different, so these two groups of tokens are left in the data.

An initial logistic regression, conducted to see the influence of the independent variables on the dependent variable, reveals that the entire model is significant with a probability of less than .0001, or a less than 1 in 1,000 chance the results are due to chance. This multivariate analysis reveals that all of the factor groups with the exception of gender and word frequency significantly condition production of the monophthong.

Birthplace is significant at a probability (p) less than .0001 and birthdate at a probability equal to .0275. The probabilities for the significant internal factor groups are all less than .0001. The reported R Square in this initial logistic regression is .22; in other words, the factor groups analyzed in this logistic regression account for 22% of the variation found in the data set.

Because the factor group birthplace contains the four factors Columbus, Bullock County, Pike County, and Houston County, bivariate analysis is conducted to uncover patterns within those four individual factors. Factors that exhibit the monophthong at statistically similar rates are combined into a single factor in order to improve the fit of the data to the model. As can be seen in Table 4, speakers from Pike County are the most monophthongal at 95.98 percent, producing 1,002 monophthongs of 1,044 opportunities to do so. Houston County speakers are also very monophthongal at 92.78 percent, though less so than the Pike County speakers. Speakers from Bullock County are slightly more monophthongal, 90.37 percent, than Columbus-born speakers are, 89.01 percent.

Table 4: All Data, Birthplace

Count Row %	a:	ai	
Pike County	1002 95.98%	42 4.02%	1044
Houston County	2492 92.78%	194 7.22%	2686
Bullock County	657 90.37%	70 9.63%	727
Columbus	591 89.01%	73 10.99%	664
	4742	379	5121

Significant at p less than .0001

Pairwise Chi Square tests reveal that all three Alabama areas are significantly different from each other, consistent with Pederson et al.'s (1986) dialectology. Columbus and Bullock County speakers produce the monophthong at statistically similar rates, consistent with Pederson et al. but counter to Carver (1987). Because Columbus and Bullock County are statistically similar to each other and are both in the same dialect area according to Pederson et al., these two factors are recoded into one. Results are displayed in Table 5.

Table 5: All Data, Birthplace, Recoded

Count Row %	a:	ai	
Pike County	1002 95.98%	42 4.02%	1044
Houston County	2492 92.78%	194 7.22%	2686
Bullock County and Columbus	1248 89.72%	143 10.28%	1391
	4742	379	5121

Significant at p less than .0001

With respect to birthdate, Table 6 shows that the oldest speakers produce the monophthong at the highest rate. The remaining speakers produce the monophthong at lower rates, consistent with previous research (Bailey & Bernstein, 1989; Bowie, 2001; Crane, 1977; Edwards, 1997; Fridland, 2003; Head, 2003; McNair, 2005).

Table 6: All Data, Birthdate

Count Row %	a:	ai	
1896	311 98.11%	6 1.89%	317
1906	60 96.77%	2 3.23%	62
1898	138 94.52%	8 5.48%	146
1927	745 93.95%	48 6.05%	793
1928	88 92.63%	7 7.37%	95
1912	622 92.28%	52 7.72%	674
1919	658 92.16%	56 7.84%	714
1915	1217 91.92%	107 8.08%	1324
1923	246 91.45%	23 8.55%	269
1916	412 91.15%	40 8.85%	452
1935	245 89.09%	30 10.91%	275
	4742	379	5121

Significant at $p = .0004$

Speakers born in 1912, 1915, 1916, 1919, and 1923 produce the monophthong at what appear to be very similar rates; pairwise testing verifies this. Thus, these factors are recoded into one, shown in Table 7.

Table 7: All Data, Birthdate, Recoded

Count Row %	a:	ai	
1896	311 98.11%	6 1.89%	317
1906	60 96.77%	2 3.23%	62
1898	138 94.52%	8 5.48%	146
1927	745 93.95%	48 6.05%	793
1928	88 96.63%	7 7.37%	95
1912 – 1923	3155 91.90%	278 8.10%	3433
1935	245 89.09%	30 10.91%	275
	4742	379	5121

Significant at p less than .0001

Further pairwise Chi Square testing of the factors reveals that speakers born in 1927 and 1928 produce the monophthong at a statistically similar rate to those born between 1912 and 1923, and thus these factors are combined. Chi Square testing of the oldest ages reveals that the speakers born in 1898 and 1906 produce the monophthong at a statistically similar rate, but that the speaker born in 1896 does not. Thus, only the 1898 and 1906 factors are combined. The recoded factors are displayed below in Table 8.

Table 8: All Data, Birthdate, Recoded

Count Row %	a:	ai	
1896	311 98.11%	6 1.89%	317
1898 – 1906	198 95.19%	10 4.81%	208
1912 – 1928	3988 92.29%	333 7.71	4321
1935	245 89.09%	30 10.91%	275
	4742	379	5121

Significant at p less than .0001

Further pairwise Chi Square testing reveals that the 1898 to 1906 factor is statistically similar to the 1912 to 1928 factor. This factor is also statistically similar to the 1935 factor, but the 1898 to 1906 factor is statistically different from the 1935 factor. Thus, the two factors 1898 to 1906 and 1912 to 1928 are recoded into one. The data indicate a very high rate of production of the monophthong with speakers decreasing in rate of production over time, consistent with prior research (Bailey & Bernstein, 1989; Bowie, 2001; Crane, 1977; Edwards, 1997; Fridland, 2003; Head, 2003; McNair, 2005). The final recoded analysis is shown in Table 9.

Table 9: All Data, Birthdate, Recoded, Final

Count Row %	a:	ai:	
1896	311 98.11%	6 1.89%	317
1898 – 1928	4186 92.43%	343 7.57%	4529
1935	245 89.09%	30 10.91%	275
	4742	379	5121

Significant at less than .0001

As shown in Table 10, males are more likely to produce the monophthong than females are, at 93.39 percent and 89.62 percent respectively. This finding is consistent with previous research regarding the relationship between gender and production of the monophthong (i.e., Edwards, 1997; Fridland, 2003).

Table 10: All Data, Gender

Count Row %	a:	ai	
males	3791 93.37%	269 6.63%	4060
females	951 89.63%	110 10.37%	1061
	4742	379	5121

Significant at p less than .0001

With respect to following phonological environment, no phonological environment disfavors the monophthong, but voiceless obstruents exhibit the monophthong at the lowest rate, 70.26 percent, as shown in Table 11. Initial findings suggest the following implicational hierarchy: voiced obstruents > nasals > word boundary > vowels > liquids > voiceless obstruents. Bowie (2001), Eckert (1996), and Hazen (2000) report that liquids exhibit the highest rate of the monophthong (see Table 1 above), in contrast to the results of this study. However, consistent with previous studies, the voiceless obstruents factor exhibits the monophthong at the lowest rate; this rate is statistically significantly different from the remaining factors (i.e., Anderson, 1999, 2002; Bowie, 2001; Eckert 1996). As in Head (2003), word boundary exhibits the monophthong at the highest rate. The remaining ordering is different from that reported in her study.

Table 11: All Data, Following Phonological Environment

Count Row %	a:	ai	
voiced obstruents	492 97.81%	11 2.19%	503
nasals	651 97.02%	20 2.98%	671
word boundary	2822 96.08%	115 3.92%	2937
vowels	32 94.12%	2 5.88%	34
liquids	237 93.68%	16 6.32%	253
voiceless obstruents	508 70.26%	215 29.82%	723
	4742	379	5121

Significant at p less than .0001

Pairwise testing reveals no clear pattern of possible combinations among the factors, so they are left separate. However, the word boundary factor (N = 2939), has more than four times the number of tokens as the next largest factor, voiceless obstruents (N = 721); 84.5 percent of the tokens in the word boundary factor are *I* (N = 2484). This suggests that inclusion of *I* may be skewing results; analysis without these tokens is discussed below.

The analysis suggests additional frequency effects even though frequency is not identified as a significant factor group in the initial logistic regression. As discussed in the literature review, word frequency affects the diffusion of sound changes throughout the lexicon, and thus the effects of word frequency on the rate of production of the monophthong are worth investigating here. Bybee (2000, 2007) and Phillips (1998, 2000) both report that higher-frequency words are affected by sound change before lower-frequency words. This study investigates a production variable. Since sound changes begin as production variables, decreasing rates of the monophthong should be associated with decreasing frequency (Bybee, 2007; Phillips, 1998, 2000). Bivariate analysis is conducted to uncover patterns in the data.

Initial coding operationalizes frequency as a continuous variable; that is, each word is associated with its own frequency. A Chi Square analysis of the factor group frequency indicates that, apart from other factor groups, frequency is significant. In order to improve the fit of the data to the model, this factor group is examined for the ranges in which frequencies tend to group. There are breaks in the data at 1,200, 5,000, and 10,000, so frequency is recoded as four mutually exclusive factors: 0 to 1,199; 1,200 to 4,999; 5,000 to 9,999; and greater than 10,000. With this recoding, frequency is significant in the logistic regression at a probability equal to .0026.

All four frequency factors favor production of the monophthong, but results, as shown in Table 12, indicate the following hierarchy: greater than 10,000 > 0 to 1,199 > 1,200 to 4,999 > 5,000 to 9,999. In light of previous research (Bybee, 2007; Phillips, 1998, 2000), the ordering of frequency factors in this study is unexpected. Like the effect of word boundary, the high number of *I* tokens (N = 2484) in the fourth quartile may be skewing the distribution of results.

Table 12: All Data, Frequency

Count Row %	a:	ai	
greater than 10,000	2441 96.06%	100 3.94%	2541
0 – 1,199	963 93.22%	70 6.78%	810
1,200 – 4,999	735 90.74%	75 9.26%	810
5,000 – 9,999	603 81.82%	134 18.18%	736
	4742	379	5121

Significant at p less than .0001

Word class is significant in the initial logistic regression at p less than .0001. As can be seen in Table 13, no word class disfavors the monophthong; however, content words, such as adjectives, adverbs, nouns, and verbs, are more likely to trigger a monophthong than function words, such as conjunctions and prepositions. Pronouns (i.e., *mine, my, I'll, I've, I'd, I'm*), the factor with the highest number of monophthongs, are the exception to this tendency. These results, which are consistent with previous research (Bowie, 2001; Head, 2003), also identify *I* as a potential curiosity.

Table 13: All Data, Word Class

Count Row %	a:	ai	
pronouns	2884 96.17%	115 3.83%	3001
nouns	944 94.49%	55 5.51%	999
adverbs	110 90.16%	12 9.84%	122
verbs	344 89.82%	39 10.18%	383
adjectives	285 81.90%	63 18.10%	348
prepositions	107 66.46%	54 33.54%	161
conjunctions	66 61.68%	41 38.32%	122
	4742	379	5121

Significant at p less than .0001

Pairwise testing reveals that conjunctions and prepositions, both function words, are statistically similar, and thus they are combined into one factor, shown in Table 14.

Table 14: All Data, Word Class, Recoded

Count Row %	a:	ai	
pronouns	2886 96.17%	115 3.83%	3001
nouns	944 94.49%	55 5.51%	999
adverbs	110 90.16%	12 9.84%	122
verbs	344 89.82%	39 10.18%	383
adjectives	285 81.90%	6-3 18.10%	348
conjunctions and prepositions	173 64.55%	95 35.44%	268
	4742	379	5121

Significant at p less than .0001

Similar to Head (2003), this study finds that pronouns are the most monophthongal, while prepositions and conjunctions are the least so. Head (2003) reports no statistically significant differences among the remaining word classes, but this study finds the remaining word classes to be statistically different from each other. Bowie (2001) reports that nouns, adverbs, and verbs favor the monophthong the most in his study, similar to the findings of this study.

As function words, the pronouns might be expected to pattern more similarly to conjunctions and prepositions. However, the pronoun factor, like the word boundary factor in analysis of following phonological environment, contains 2,484 tokens of *I*, which represent 82.77 percent of the pronoun factor. Tokens of *I* comprise 48.51 percent of the data as a whole. This suggests that the high number of *I* tokens in comparison to other tokens in this corpus is affecting results for this factor.

Stress is a significant factor group in the initial logistic regression (p less than .0001). As can be seen in Table 15, unstressed words are significantly more likely to be monophthongal than stressed words are. This finding contrasts with Head's (2003) finding that stress does not influence production of the monophthong.

Table 15: All Data, Stress

Count Row %	a:	ai	
unstressed	513 98.84%	6 1.16%	519
stressed	4229 91.89%	373 8.11%	4602
	4742	379	5121

Significant at p less than .0001

As discussed above, to improve the fit of the data to the model, factors within factor groups that Chi Square testing identifies as statistically similar have been combined within individual factor groups. Then a final logistic regression is conducted to verify initial results and retest the hypothesis. The final logistic regression of this data set reveals that, of the external factor groups, only birthplace (p less than .0001) and birthdate (p = .0004) condition the alternation between the monophthong and the diphthong. The significance of the birthdate factor group improves from the initial logistic regression (p = .0275). All the internal factor groups are significant in the final analysis, with following phonological environment significant at p less than .0001, stress at p less than .0001, frequency at .0023, and word class at .0365. Significance decreases between the two logistic regressions for the factor groups frequency and word class. The R Square reveals a slight improvement in the fit of the data to the model, with the recoded factor groups now accounting for 22.12 percent of the variation within the data.

At first glance, most of the findings in this analysis are consistent with previous research. However, closer examination reveals intriguing differences, possibly due to the behavior of the *I* tokens, which represent 48.51 percent of the data set. Thus, additional analysis is undertaken below to investigate this possibility.

Analysis of the I Data Set

The treatment of *I* tokens is not often discussed in previous research. However, one study limits tokens of *I* in the data set due to their high frequency (Bowie, 2001). Alternatively, the tokens are excluded due to their near categorical nature (Head, 2003) or their high frequency (Eckert, 1996). In this study, the *I* tokens are not categorical or near categorical, exhibiting the monophthong at 96.01 percent, but they are highly frequent in this data set**.** Their frequency may be affecting results. Therefore, the large number of I tokens provides an opportunity to analyze them separately from the remaining data.

The contractions containing *I* (*I'm, I'll, I've,* and *I'd*) are also considered for exclusion from the full data set since they exhibit a similarly high rate of production of the monophthong (96.00 percent). Head (2003) also excludes these tokens due to their nearly categorical nature. However, Neu (1980) excludes only one nearly categorical type of token, *and*, in her study because the other, *just,* does not skew results. In this study, the contractions containing *I* are not such a high percentage of the total data as the tokens *I*. Therefore, the rates of production of the monophthong of the full data set and of the data without the contractions are compared. The rates are statistically similar for the two data sets. Thus, the contractions remain.

The overall rate of production of the monophthong for the lexical item *I* is 96.01 percent, higher than the rate of production of the monophthong in the entire data set (92.60 percent). An initial logistic regression of the *I* tokens reveals that only birthplace, birthdate, and stress significantly condition the alternation between the monophthong and the diphthong, at a p less than .0001 for each factor group. The R Square reported for this data is 9 percent.

Bivariate analysis of the birthplace factor group, as can be seen in Table 16, indicates that speakers from Pike County produce the monophthong most often, at 98.71 percent. Pike County speakers are the most monophthongal in the larger data set as well, but otherwise the rank ordering of birthplace differs between the two data sets.

Table 16: *I* only, Birthplace

Count Row %	a:	ai	
Pike County	460 98.71%	6 1.29%	466
Columbus	319 97.26%	9 2.74%	328
Houston County	1226 95.41%	59 4.59%	1285
Bullock County	380 93.83%	25 6.17%	405
	2385	99	2484

Significant at p = .0009

Unlike in the larger data set, Pike County and Columbus speakers are statistically similar to each other, and so are recoded into one factor for this analysis, shown in Table 17.

Table 17: *I* Only, Birthplace, Recoded

Count Row %	a:	ai	
Columbus and Pike County	779 98.11%	15 1.89%	794
Houston County	1226 95.41%	59 4.59%	466
Bullock County	380 93.83%	25 6.17%	2484
	2385	99	2484

Significant at p = .0004

In the initial logistic regression, gender is not significant (p = .7468). However, males do produce significantly more [a:] when the factor group is examined in isolation from other factor groups, as can be seen in Table 18. This finding is consistent with the larger data set as well as previous research (i.e., Edwards, 1997; Fridland, 2003).

Table 18: *I* Only, Gender

Count Row %	a:	ai	
males	1832 96.47%	67 3.53%	1899
females	553 94.53%	32 5.47%	585
	2385	99	2484

Significant at p = .0358

Like gender, birthdate is not significant in the initial logistic regression. Bivariate analysis is conducted to see if the data clusters into groups, thereby improving the fit of the data to the model. Table 19 shows the initial birthdate data.

Table 19: *I* Only, Birthdate

Count Row %	a:	ai	
1896	173 100%	0 0%	173
1898	67 100%	0 0.00%	67
1906	30 100%	0 0.00%	30
1928	49 100%	0 0.00%	49
1912	273 98.20%	5 1.80%	278
1927	390 97.50%	10 2.50%	400
1916	238 96.36%	9 3.64%	247
1915	557 96.03%	23 3.97%	580
1923	120 96.00%	5 4.00%	125
1919	346 91.78%	31 8.22%	377
1935	142 89.87%	16 10.13%	158
	2385	99	2484

Significant at p less than .0001

The factors 1927 and 1928 are compared using a Fisher's Exact Test[3]; these two factors exhibit the monophthong at statistically similar rates and are recoded into a single factor. Speakers born in 1915 and 1916 produced the monophthong in this smaller data set at essentially the same rates. Because the speakers born between 1896 and 1906 produce the monophthong categorically, they are combined. Recoded factors are shown in Table 20.

Table 20: *I* Only, Birthdate, Recoded

Count Row %	a:	ai	
1896 – 1906	270 100%	0 0%	270
1928	49 100%	0 0.00%	49
1912	273 98.20%	5 1.80%	278
1923 – 1927	510 97.14%	15 2.86%	525
1915 – 1916	795 96.13%	32 3.87%	827
1919	346 91.78%	31 8.22%	377
1935	142 89.87%	16 10.13%	158
	2385	99	2484

Significant at p less than .0001

Further pairwise Chi Square testing reveals that the speaker born in 1923 produces the monophthong at a rate statistically similar to speakers in the 1927 to 1928 factor; these are recoded into a single factor. Other factors that exhibit the monophthong at statistically similar rates are not combined because doing so decreases significance of the factor group. Final results are displayed in Table 21.

[3] Chi Square is not recommended when expected frequencies are five or fewer, as is the case in the above pairwise comparison. The Fisher's Exact Test is suggested when this occurs; results are reported as probabilities, and alpha is set at .05.

Table 21: *I* Only, Birthdate, Recoded, Final

Count Row %	a:	ai	
1896 – 1906	270 100%	0 0%	270
1923 – 1928	559 97.39%	15 2.61%	574
1912 – 1916	1068 96.65%	37 3.35%	1105
1919	346 91.78	31 8.22%	377
1935	142 89.87%	16 10.13%	158
	2385	99	2484

Significant at p less than .0001

As in the larger data set and consistent with prior research, the oldest speakers produce the monophthong at the highest rate, and the youngest speaker produces the monophthong at lower rates (i.e., Bailey & Bernstein, 1989; Bowie, 2001; Edwards, 1997). The two exceptions are the speakers born in 1912 and 1919.

Like unstressed tokens in the larger data set, unstressed words in this data set are almost categorically monophthongal at 99.57 percent, shown in Table 22.

Table 22: *I* Only, Stress

Count Row %	a:	ai	
unstressed	460 99.57%	2 0.43%	462
stressed	1925 95.20%	97 4.80%	2022
	2385	99	2484

Significant at p less than .0001

A second logistic regression, run after recoding the birthdate and birthplace factor groups, reveals that the four factor groups considered in this analysis significantly condition the alternation between the monophthong and the diphthong. Birthplace is significant with a probability of .0140, gender is significant with a probability of .0231, birthdate is significant with a probability of less than .0001, and stress is significant with a probability of less than .0001. The R Square of this data set is 10 percent, considerably smaller than the R Square of the larger data set but higher than the R Square reported initially for this smaller data set.

Analysis of Data Excluding I Tokens

After exclusion of *I* tokens, 2,637 tokens remain for analysis. In the data set excluding *I* tokens, the frequency of [a:] is 89.38 percent, lower than in the full data set.

The initial logistic regression of the reduced data set reveals birthplace ($p = .0000$)[4], gender ($p = .0232$), following phonological environment ($p = .0000$), and word class ($p = .0000$) to significantly condition production of the monophthong. Frequency divided into quartiles is significant at $p = .0020$. The reported R Square is 29 percent, higher than that reported for the initial analysis.

As can be seen in Table 23, Pike County speakers are the most monophthongal at 93.77 percent, and Columbus-born speakers the least monophthongal at 80.95 percent, consistent with analysis of the full data set.

---

[4] JMP only reports probabilities to the fourth decimal point; the reported probability is small enough that JMP rounds the number to zero.

Table 23: *I* Excluded, Birthplace

Count Row %	a:	ai	
Pike County	542 93.77%	36 6.23%	578
Houston County	1266 90.36%	135 9.64%	1401
Bullock County	277 86.02%	45 13.98%	322
Columbus	272 80.95%	64 19.05%	336
	2357	280	2637

Significant at p less than .0001

Pairwise tests reveal that Columbus and Bullock County speakers produce the monophthong at statistically similar rates. Pike County and Houston County speakers are statistically different from each other as well as from Columbus and Bullock County speakers. Thus, the factors of Columbus and Bullock County are combined into one, as shown in Table 24. This result patterns similarly to analysis of birthplace in the full data set.

Table 24: *I* Excluded, Birthplace, Recoded

Count Row %	a:	ai	
Pike County	542 93.77%	36 6.23%	578
Houston County	1266 90.36%	135 9.64%	1401
Columbus and Bullock County	549 83.43%	109 16.57%	658
	2357	280	2637

Significant at p less than .0001

As can be seen in Table 25, males are significantly more likely to produce the monophthong than females, at 90.65 to 83.61 percent. This result patterns similarly to the previous analyses of both the full data set and the *I* tokens as well as to prior research (i.e., Edwards, 1997; Fridland, 2003).

Table 25: *I* Excluded, Gender

Count Row %	a:	ai	
males	1959 90.65%	202 9.35%	2161
females	398 83.61%	78 16.39%	476
	2357	280	2637

Significant at p less than .0001

Birthdate, as a continuous variable, is not significant in the initial logistic regression, at p = .0555. The data is shown below in Table 26. As in earlier analyses of all data and *I* only, birthdate is examined more closely for clusters within the data in an attempt to improve the fit of the data to the model.

Table 26: *I* Excluded, Birthdate

Count Row %	a:	ai	
1896	138 95.83%	6 4.17%	144
1906	30 93.75%	2 6.25%	32
1919	312 92.58%	25 7.42%	337
1927	355 90.33%	38 9.67%	393
1898	71 89.87%	8 10.13%	79
1915	660 88.71%	84 11.29%	744
1912	349 88.13%	47 11.87%	396
1935	103 88.03%	14 11.97%	117
1923	126 87.50%	18 12.50%	144
1916	174 84.88%	31 15.12%	205
1928	39 84.78%	7 15.22%	46
	2357	280	2637

Significant at p = .0319

To see if the birthdate data clusters, factors with similar rates of production of the monophthong and that are consecutive are examined for statistical similarity. A pairwise test is first conducted on the birthdates 1912 and 1915 since these speakers produce the monophthong at such similar rates, at 88.13 percent and 88.71 percent respectively. The test reveals these two factors to be statistically similar, so they are recoded into one. In addition, pairwise testing also reveals speakers born in 1919, 1923, and 1927 to produce the monophthong at statistically similar rates, and thus these factors are recoded into one as well, as can be seen in Table 27.

Table 27: *I* Excluded, Birthdate, Recoded

Count Row %	a:	ai	
1896	138 95.83%	6 4.17%	144
1906	30 93.75%	2 6.25%	32
1919 – 1927	793 90.73%	81 9.27%	874
1898	71 89.87%	8 10.13%	79
1912 – 1915	1009 88.51%	131 11.49%	1140
1935	103 88.03%	14 11.97%	117
1916	174 84.88%	31 15.12%	205
1928	39 84.78%	7 15.22%	46
	2357	280	2637

Significant at p = .0211

Speakers born in 1928 and 1935 produce the monophthong at statistically similar rates and are recoded into a single factor. Rates of production of the monophthong of speakers born in 1896, 1898, and 1906 are also tested; these speakers produce the monophthong at rates statistically similar to each other and thus are also recoded into one factor. Recoding improves fit of the data to the model, shown in Table 28.

Table 28: *I* Excluded, Birthdate, Recoded

Count Row %	a:	ai	
1896 – 1906	239 93.73	16 6.27	255
1919 - 1927	793 90.73%	81 9.27%	874
1912 – 1915	1009 88.51%	131 11.49%	1140
1916	174 84.88%	31 15.12%	205
1928 – 1935	142 87.12	21 12.88	163
	2357	280	2637

Significant at p = .0101

In further testing, the rate of production of the monophthong of the speaker born in 1916 is first compared to the rate produced by speakers in the 1912 and 1915 factor, and then the 1919 to 1927 factor. The speaker born in 1916 patterns statistically similarly to speakers in the 1912 and 1915 factor but not the 1919 to 1927 factor. Thus, the 1916 factor is combined with the 1912 to 1915 factor. Testing also reveals speakers in the 1928 to 1935 factor to pattern statistically similarly to speakers in the 1919 to 1927 factor, and so these two factors are recoded into one. However, speakers in the 1896 to 1906 factor produce the monophthong at a rate statistically different from those in the 1912 to 1916 factor, and so they are not combined. Finally, though testing does show the 1912 to 1916 factor to be statistically similar to the 1919 to 1935 factor, combination of these two factor does not improve the fit of the data to the model; the probability of error increases. The final recoding of the birthdate factor group is displayed in Table 29 below. The rate of production of the monophthong is highest for the oldest speakers, consistent with prior research, but the relationship between age and production of the monophthong is not straightforward; the youngest speakers produce the monophthong at a higher rate than those born between 1912 and 1916 (i.e., Bailey & Bernstein, 1989, Bowie, 2001; Edwards, 1997; Fridland, 2003).

Table 29: *I* Excluded, Birthdate, Recoded, Final

Count Row %	a:	ai	
1896 – 1906	239 93.73%	16 6.27%	255
1919 – 1935	935 90.16%	102 9.84%	1037
1912 – 1916	1183 89.96%	162 11.04%	1345
	2357	280	2637

Significant at p = .0095

In regard to following phonological environment, results from bivariate analysis suggest the same implicational hierarchy as the full data set, shown in Table 10: voiced obstruents > nasals > word boundary > vowels > liquids > voiceless obstruents. Also, as in analysis of the full data set, no environment disfavors production of the monophthong, and the differences in rate of production of the monophthong between the voiceless obstruents and the remaining factors are statistically significant. Though voiced obstruents and nasals exhibit the monophthong at statistically similar rates, as do vowels and liquids, the probability of error increases when these factors are combined. Table 30 displays the results for this factor group.

Table 30: *I* Excluded, Following Phonological Environment

Count Row %	a:	ai	
voiced obstruents	492 97.81%	11 2.19%	503
nasals	650 97.01%	20 2.99%	670
word boundary	437 96.47%	16 3.53%	453
vowels	32 94.12%	2 5.88%	34
liquids	238 93.70%	16 6.30%	254
voiceless obstruents	508 70.26%	215 29.74%	723
	2357	280	2637

Significant at p less than .0001

The ranking of the frequency quartiles when the tokens of *I* are excluded is shown in Table 31. In this smaller data set, the rate of production of the monophthong is higher for the most frequent words, but the rank ordering remains the same as in analysis of the full data set (see Table 11). As the literature on frequency effects suggests, the most frequent words exhibit the monophthong at the highest rate, but the ordering of the remaining factors is not consistent with that research (Bybee, 2007; Phillips, 1998, 2000).

Table 31: *I* Excluded, Frequency

Count Row %	a:	ai	
greater than 10,000	56 98.25%	1 1.75%	57
0 – 1,199	963 93.22%	70 6.78%	1033
1,200 – 4,999	735 90.74%	75 9.26%	810
5,000 – 9,999	603 81.82%	134 18.18%	737
	2357	280	2637

Significant at p less than .0001

With respect to word class, as in analysis of the full data set, no factor disfavors the monophthong. Content words exhibit the monophthong at a higher rate than do function words, with the exception of pronouns. As in analysis of the full data set, even with tokens of *I* excluded, pronouns are more monophthongal than the remaining factors. Bivariate results for tokens with *I* excluded are displayed in Table 32 below.

Table 32: *I* Excluded, Word Class

Count Row %	a:	ai	
pronouns	501 96.91%	16 3.16%	517
nouns	944 94.49%	55 5.51%	999
adverbs	110 90.16%	12 9.84%	122
verbs	343 89.79%	39 10.21%	382
adjectives	285 81.90%	63 18.10%	348
prepositions	108 66.67%	54 33.33%	162
conjunctions	66 61.68%	41 38.32%	107
	2357	280	2637

Significant at p less than .0001

As in the full data set, conjunctions and prepositions are statistically similar and so are recoded into a single factor, shown below in Table 33. Analysis results in an implicational hierarchy the same as the one indicated in analysis of the full data set (see Table 12): pronouns > nouns > adverbs > verbs > adjectives > prepositions and conjunctions.

Table 33: *I* Excluded, Word Class, Recoded

Count Row %	a:	ai	
pronouns	501 96.91%	16 3.16%	517
nouns	944 94.49%	55 5.51%	999
adverbs	110 90.16%	12 9.84%	122
verbs	343 89.79%	39 10.21%	382
adjectives	285 81.90%	63 18.10%	348
prepositions and conjunctions	174 64.68%	95 35.32%	269
	2357	280	2637

Significant at p less than .0001

While unstressed tokens are more likely to be monophthongal than stressed tokens in the full data set as well as the smaller one analyzed here, stress does not significantly condition the alternation in either the logistic regression or bivariate analysis of the data excluding *I* tokens. This result is similar to what Head (2003), who excluded *I* tokens, reports for speakers in Elba. Stress as a conditioning factor seems to be due primarily to the *I* tokens, which are excluded from (Eckert, 1996; Head, 2003) or limited in analysis (Bowie, 2001) when discussed.

A final logistic regression of the data set with the *I* tokens excluded reveals that, of the external factor groups, only birthplace (p = .0000) and birthdate (p = .0204) condition the alternation between the monophthong and the diphthong in this data set. The internal factor groups following phonological environment (p = .0000), frequency (p = .0042), and word class (p=.0007) also significantly condition the alternation. Gender and stress are not identified as significantly conditioning factor groups. Recoding of the factor groups reduces the R Square from 29 to 27 percent.

So far, findings from this study are consistent with findings from previous research except in the case of frequency. It seems odd that the least frequent words (0 to 1,299) are more monophthongal than all but the most frequent words (greater than 10,000). Thus, the frequency groupings are examined more closely in the next section in order to determine what might be affecting results.

Exclusion and Analysis of Tokens of Like, Liked, Alike and Unlike

Analysis of the full data set suggests that inclusion of the large number of *I* tokens may be affecting frequency results. However, the factor group frequency in the smaller data set without the *I* tokens is also inconsistent with the literature on frequency effects, so the corpus is again considered. The interviews analyzed for this study, which focus on the textile industry, contain forty-five machine and industry-specific words, including words containing (ai) such as *winders*, *dyehouse*, and *textile*, but *The American Heritage Word Frequency Book* includes very few of these terms. The textile-specific terms are present at a relatively higher frequency in the corpus. Removal of textile terminology from the data set does slightly change the rate of the production of the monophthong in the 0 to 1,200 quartile from 93.22 percent to 93.50 percent but does not change the rank ordering of the quartiles.

Since the textile-specific terms in the lowest quartile do not affect the ordering of the frequency results, the third quartile (5,000 to 9,999) is examined next. The monophthong appears at a percentage of 82.34 in this quartile, considerably lower than the other frequency quartiles. Displayed below in Table 34, the 5,000 to 9,999 factor (N=737) includes 273 tokens of *my,* 229 tokens of *time,* 210 tokens of *like,* 18 tokens of *find,* and 7 tokens of *write. My, time,* and *find* are in phonological and word class factors highly correlated with the monophthong. *Write* and *like*, members of the voiceless obstruents factor, are not as highly correlated with the monophthong.

As can be seen in Table 34, *write, time, my,* and *find* are produced as a monophthong in more than 90 percent of possible occurrences, similar to other tokens in the full data set. However, *like,* which grammatically can be a verb, preposition, conjunction, adjective, noun, or adverb, exhibits unusual behavior. Unlike the rest of the data in this study, tokens of *like* disfavor the monophthong, exhibiting the variable at 42.86 percent in this quartile. Without these tokens, this factor exhibits the monophthong at a rate of 97.50 percent. A member of the following voiceless obstruents factor, *like* is likely affecting the rate of the monophthong in this frequency quartile.

Table 34: Distribution of Tokens in the 5,000 – 9,999 Frequency Quartile

Count Row %	a:	ai	
write	7 100%	0 0%	7
time	226 98.70%	3 1.30%	229
my	264 96.70%	9 3.30%	273
find	17 94.44%	1 5.56%	18
like	90 42.86%	120 57.14%	210
	603	134	737

As discussed in the literature review, Neu (1980) argues for exclusion of categorical or nearly categorical tokens when they behave differently from the remaining data set. In this case, although not categorical, *like* is affecting data results atypically. Therefore, 210 tokens of *like,* 10 of *liked,* 1 of *unlike*, and 1 of *alike*, a total of 222 tokens, are analyzed separately from the data set.

In the 222 tokens of *like, liked, unlike,* and *alike* in the remaining data, the monophthong appears at a rate of 41.89 percent, a slightly lower rate than that of tokens of *like* only. An initial logistic regression of tokens containing the root *like* indicates that the factor groups together account for nearly 18 percent of the variation in this small data set. However, no factor group in the initial logistic regression significantly conditions the alternation between the monophthong and the diphthong. Bivariate analysis is undertaken below to see if, as in previous analyses, patterns reveal themselves when individual factor groups are analyzed separately from each other.

As can be seen in Table 35, speakers born in Pike County are the only ones to produce the monophthong more than half the time. For the remaining speakers, the monophthong is disfavored in words containing the root *like.*

Table 35: *Like* Only, Birthplace

Count Row %	a:	ai	
Pike County	36 66.67%	18 33.33%	54
Houston County	40 38.10%	65 61.90%	105
Bullock County	12 37.50%	20 62.50%	32
Columbus	5 16.13%	26 83.87%	31
	93	129	222

Significant at p less than .0001

Houston County and Bullock County speakers produce the monophthong at a statistically similar rate, and therefore these two factors are recoded into one, as can be seen in Table 36.

Table 36: *Like* Only, Birthplace, Recoded

Token	a:	ai	
Pike County	36 66.67%	18 33.33%	54
Houston and Bullock Counties	52 37.96%	85 62.04%	137
Columbus	5 16.13	26 83.87%	31
	93	129	222

Significant at p less than .0001

With respect to gender, males are more likely to produce the monophthong, as in previous analyses. However, this relationship is not significant (p = .0600) in isolation from other factor groups, as can be seen in Table 37.

Table 37: *Like* Only, Gender

Token	a:	ai	
males	77 45.29%	93 54.71%	170
females	16 30.77%	36 69.23%	52
	93	129	222

Not significant, p = .0600

Birthdate is not significant in the initial logistic regression of this small data set, but is significant when considered in isolation from other factor groups, as shown below in Table 38. As in previous analyses, individual factors are tested for statistical similarities in an effort to increase cell size and improve the fit of the data to the model.

Table 38: *Like* Only, Birthdate

Count Row %	a	ai	
1896	7 87.50%	1 12.50%	8
1912	21 56.76%	16 43.24%	37
1927	19 47.50%	21 52.50%	40
1919	9 45.00%	11 55.00%	20
1916	10 4000%	15 60.00%	25
1915	21 32.81%	43 67.19%	64
1935	2 28.57%	5 71.43%	7
1898	1 25.00%	3 75.00%	4
1923	3 23.08%	10 76.92%	13
1928	0 0.00%	4 100.00%	4
	93	129	222

Significant at p = .0136

Chi Square is valid for testing statistical similarity when the expected cell frequencies are more than five. When expected frequencies are fewer than five, a Fisher's Exact test is conducted. Testing reveals that speakers born in 1915, 1916, and 1919 produce the monophthong at statistically similar rates, and so these three factors are recoded into one. The speaker born in 1906 produces no tokens containing the root *like* and so is excluded from analysis. No other factors are statistically similar. With recoding, the relationship between birthdate and rate of production of the monophthong strengthens, shown below in Table 39. Previous research has reported a general tendency for younger speakers to produce the monophthong at a lower rate than older speakers do;

results reported here are generally consistent with that but for three speakers born in 1898, 1927, and 1935. However, research also reports that rate of production among tokens in the voiceless obstruent factor is increasing among younger speakers. This is inconsistent with results reported for tokens containing the root *like* (Bailey and Bernstein, 1989; Bowie, 2001; Crane, 1977; Edwards, 1997; Fridland, 2003; Head, 2003; McNair, 2005).

Table 39: *Like* Only, Birthdate, Recoded

Count Row %	a	ai	
1896	7 87.50%	1 12.50%	8
1912	21 56.76%	16 43.24%	37
1927	19 47.50%	21 52.50%	40
1915 – 1919	40 36.70%	69 63.30%	109
1935	2 28.57%	5 71.43%	7
1898	1 25.00%	3 75.00%	4
1923	3 23.08%	10 76.92%	13
1928	0 0.00%	4 100.00%	4
	93	129	222

Significant at p = .0063

Of the internal factor groups, following phonological environment is a constant in this data set and so cannot be analyzed, but frequency, word class, and stress are analyzed. Of the four frequency quartiles, only two quartiles appear in the *like* data set: 0 to 1,199 and 5,000 to 9,999. The 0 to 1,199 quartile contains 210 tokens of *like*; the 5,000 to 9,999 quartile contains 12 tokens of *liked, unlike,* and *alike*. As can be seen in Table

40, the rate of production of the monophthong decreases as word frequency decreases, consistent with previous research (Bybee, 2007; Phillips, 1998, 2000).

Table 40: *Like* Only, Frequency

Count Row %	a	ai	
5,000 – 9,999	91 42.52%	123 57.48%	214
0 – 1,199	2 25.00%	6 75.00%	8
	93	129	222

Not significant, p = .3096

With respect to word class, the distribution of tokens is significant in the logistic regression. As can be seen in Table 41, only when *like* is used as a conjunction (i.e., "*It seems like it was two bedrooms.*") does it favor the monophthong, at 53.57 percent.

Table 41: *Like* Only, Word Class

Count Row %	a	ai	
conjunctions	45 53.57%	39 46.43%	84
adjectives	9 37.50%	15 62.50%	24
prepositions	29 36.25%	51 63.75%	80
verbs	10 30.30%	23 69.70%	33
adverbs	0 0.00%	1 100.00%	1
	93	129	222

Not significant at p = .0647

Pairwise testing reveals that, except for conjunctions, the word classes are statistically similar to each other. Therefore, adjectives, prepositions, verbs, and adverbs are combined into one factor, as can be seen in Table 42. Recoding strengthens the relationship between word class and production of the monophthong.

Table 42: *Like* Only, Word Class, Recoded

Count Row %	a	ai	
conjunctions	45 53.57%	39 46.43%	84
adjectives, prepositions, verbs, adverbs	48 34.78%	90 65.22%	138
	129	222	93

Significant at p = .0060

As in previous analyses, stressed tokens are more likely to be produced as a monophthong than unstressed tokens are, but the factor group stress is not significant in isolation, as can be seen in Table 43.

Table 43: *Like* Only, Stress

Count Row %	a	ai	
stressed	93 42.08%	128 57.92%	221
unstressed	0 0.00%	1 100.00%	1
	129	222	93

Not significant, p = .2967

A logistic regression of the recoded factor groups reveals that birthplace (p = .0007) and word class (p = .0027) are the only factor groups that significantly condition the alternation between the monophthong and the diphthong in the data set of tokens containing only the root *like.* The reported R Square is 17 percent in the second logistic regression for this data set, smaller than that reported in the initial logistic regression.

Analysis of Data Excluding I and Words Containing the Root Like

Exclusion of tokens containing the root *like* leaves 2,415 tokens for analysis. These tokens exhibit the monophthong 93.75 percent of the time, a higher rate than in the full data set. A logistic regression indicates that only birthplace, gender, and following phonological environment condition the alternation in this smaller data set, at p = .0000, p = .0333, and p = .0006, respectively. The reported R Square indicates that 14 percent of the variation between the monophthong and the diphthong is accounted for by the factor groups analyzed in this smaller data set.

Bivariate analysis of birthplace, displayed in Table 44, indicates that Pike County speakers are the most monophthongal, at 96.56 percent, and that Columbus speakers are the least monophthongal, at 87.54 percent, consistent with prior analyses.

Table 44: *I* and the Root *Like* Excluded, Birthplace

Count Row %	a:	ai	
Pike County	506 96.56%	18 3.44%	524
Houston County	1226 94.60%	70 5.40%	1296
Bullock County	265 91.38%	25 8.62%	290
Columbus	267 87.54%	38 12.46%	305
	2264	151	2415

Significant at p less than .0001

Pairwise tests of the factors in this factor group reveal that speakers from the three Alabama locations, Pike County, Houston County, and Bullock County, produce the monophthong at statistically similar rates. Columbus speakers are statistically dissimilar to Pike County and Houston County speakers though similar to Bullock County speakers. The Pike County and Houston County factors are combined, as they are closest to each other in rate of production of the monophthong. The difference between Houston County and Bullock County speakers is nearly significant, so Houston County is not combined with the other two Alabama locations. Results are displayed below in Table 45.

Table 45: *I* and the Root *Like* Excluded, Birthplace, Recoded

Count Row %	a:	ai	
Pike and Houston Counties	1732 95.16%	88 4.84%	1820
Bullock County	265 91.38%	25 8.62%	290
Columbus	267 87.54%	38 12.46%	305
	2264	151	2415

Significant at p less than .0001

As in previous analyses and consistent with prior research (i.e., Edwards, 1997; Fridland, 2003), males produce the monophthong at a higher rate, 94.53 percent, than females do, 90.09 percent.

Table 46: *I* and the Root *Like* Excluded, Gender

Count Row %	a:	ai	
males	1882 94.53%	109 5.47%	1991
females	382 90.09%	42 9.91%	424
	2264	151	2415

Significant at p = 0.0007

Birthdate is not identified in the logistic regression as significantly conditioning the variation in this data set. Examination of the factors in this factor group reveals that speakers born in all years produce the monophthong at statistically similar rates, so no further analysis is conducted.

Of the internal factor groups investigated in this study, only following phonological environment is significant in this smaller data set that excludes *I* and words containing the root *like.* The distribution of tokens can be seen in Table 47 below. In preceding analyses (i.e., the full data set and the data set excluding *I*), voiceless obstruents are monophthongal 70.26 percent of the time; exclusion of the *like* tokens increases this to 82.83 percent. Even with this difference, the voiceless obstruents are significantly less monophthongal than the other phonological factors are, consistent with both previous analyses and previous research (Anderson, 1999, 2002; Bowie, 2001; Hazen, 2000; Eckert, 1996; Fridland, 2003; Labov and Ash, 1997).

Table 47: *I* and the Root *Like* Excluded, Following Phonological Environment

Count Row %	a:	ai	
voiced obstruents	492 97.81%	11 2.19%	503
nasals	651 97.02%	20 2.98%	671
word boundary	437 96.47%	16 3.53%	453
vowels	32 94.12%	2 5.88%	34
liquids	237 93.68%	16 6.32%	253
voiceless obstruents	415 82.83%	86 17.17%	501
	2264	151	2415

Significant at p less than .0001

Though other linguistic factor groups, including frequency, are not significant in the initial logistic regression, the distribution of tokens in the factor group frequency is re-examined to determine if the unexpected results reported earlier change with exclusion of tokens containing the root *like*. As can be seen in Table 48, the third quartile is no longer the least monophthongal. Instead, as previous research suggests it should be (Bybee, 2007; Phillips, 1998, 2000), it is the second most monophthongal.

Table 48: *I* and the Root *Like* Excluded, Frequency

Count Row %	a:	ai	
greater than 10,000	56 98.25%	1 1.75%	57
5,000 – 9,999	512 97.90%	11 2.10%	523
0 – 1,199	961 93.765	64 6.24%	1025
1,200 – 4,999	735 90.74%	75 9.26%	810
	2264	151	2415

Significant at p less than .0001

Pairwise testing reveals that the first quartile is significantly different from the others and that the other quartiles are statistically similar to each other. However, the rank ordering of the frequency quartiles is still not consistent with previous research (Bybee, 2007; Phillips, 1998, 2000), so the statistically similar factors are not combined. Instead, the frequency quartiles are inspected for tokens affecting the rank ordering below.

Exclusion of tokens containing the root *like* affects the factor group word class, also. Earlier analyses suggest an implicational hierarchy of pronouns > nouns > adverbs > verbs > adjectives > prepositions and conjunctions. However, in this data set, both prepositions and conjunctions are considerably more monophthongal than in previous analyses (prepositions: 66.67 vs. 96.34 percent, conjunctions: 61.68 vs. 91.30 percent). Verbs, adverbs, and adjectives are also more monophthongal than in previous analyses, though the change is not as great. Results are displayed below in Table 49.

Table 49: *I* and the Root *Like* Excluded, Word Class

Count Row %	a:	ai	
pronouns	501 96.91%	16 3.09%	517
prepositions	79 96.34%	3 3.66%	82
verbs	333 95.42%	16 4.58%	349
nouns	944 94.49%	55 5.51%	999
conjunctions	21 91.30%	2 8.70%	23
adverbs	110 90.91%	11 9.09%	121
adjectives	276 85.19%	48 14.81%	324
	2264	151	2415

Significant at p less than .0001

Pairwise testing reveals the two function-word factors of pronouns and prepositions to be statistically similar, so they are combined into one. Recoding is displayed in Table 50.

Table 50: *I* and the Root *Like* Excluded, Word Class, Recoded

Count Row %	a:	ai	
pronouns and prepositions	580 96.83%	19 3.17%	599
verbs	333 95.42%	16 4.58%	349
nouns	944 94.49%	55 5.51%	999
conjunctions	21 91.30%	2 8.70%	23
adverbs	110 90.91%	11 9.09%	121
adjectives	276 85.19%	48 14.81%	324
	2264	151	2415

Significant at p less than .0001

A logistic regression of the recoded factor groups indicates that following phonological environment (p = .0000), birthplace (p = .0009), and gender (p = .0333) significantly condition the alternation between the monophthong and the diphthong in the data set excluding *I* and tokens containing the root *like*. The reported R Square in the final logistic regression indicates that these factor groups account for 14 percent of the variation between the monophthong and the diphthong in this smaller data set, less than the R Square reported in the initial logistic regression.

Exclusion and Analysis of Tokens Containing the Root Right

Returning to the unexpected ordering of frequency quartiles shown in Table 48, exclusion of tokens containing the root *like* results in expected placement of the third quartile. However, the 0 to 1,199 and 1,200 to 4,999 quartiles are not ranked as expected.

The 1,200 to 4,999 factor includes 209 tokens of *right*, another grammatically versatile word. Examination of these tokens separately from the remainder of the data reveals that they are produced as a monophthong at 81.82 percent. This rate is much closer to the overall production of the monophthong than the frequency of the monophthong in words containing the root *like*. However, the rate of the monophthong in tokens containing the root *right* is lower than the rate in the full data set (92.60 percent), in the data set excluding *I* (89.38 percent), and in the data set excluding *I* and tokens containing the root *like* (93.75 percent).

All 209 tokens of *right* and 1 token of *rights* are excluded from the data set. Analysis reveals no factor group significantly conditions the alternation between the monophthong and the diphthong in this small data set.

Analysis of Data Excluding Tokens of I and Tokens Containing the Roots Like and Right

After excluding *I* and tokens containing the roots *right* and *like*, 2,204 tokens remain for analysis. This data set is the most monophthongal of all data sets examined so far, at 94.87 percent. The initial logistic regression shows that only birthplace, gender, and following phonological environment significantly condition the alternation between [a:] and [ai] in this data set, at $p = .0000$, $p = .0007$ and $p = .0015$, respectively. The factor groups account for only 14 percent of the variation.

Table 51 shows that, as in previous analyses, Pike County speakers produce the monophthong at the highest rate, 98.49 percent, and Columbus speakers produce the monophthong at the lowest rate, 90.35 percent. Exclusion of *I, like,* and *right* tokens increases the overall percentage of the monophthong for each birthplace, but rank ordering of the four birthplaces is the same as in analysis of the full data set, the data set excluding *I*, and the data set excluding *I* and *like*.

Table 51: *I* and the Roots *Like, Right* Excluded, Birthplace

Count Row %	a:	ai	
Pike County	457 98.49%	7 1.51%	464
Houston County	1151 95.05%	60 4.95%	1211
Bullock County	249 92.22%	21 7.78%	270
Columbus	234 90.35%	25 9.65%	259
	2091	113	2204

Significant at p less than .0001

In this analysis, the difference between Houston County, Bullock County, and Columbus is not significant. Therefore, these three factors are recoded into one, displayed in Table 52.

Table 52: *I* and the Roots *Like, Right* Excluded, Birthplace, Recoded

Count Row %	a:	ai	
Pike County	457 98.49%	7 1.51%	464
Houston and Bullock Counties, Columbus	1634 93.91%	106 6.09 %	1481
	2091	113	2204

Significant at p less than .0001

As in previous analyses, males are significantly more likely to produce the monophthong than females are, at 95.69 percent to 91.16 percent, shown in Table 53.

Table 53: *I,* and the Roots *Like, Right* Excluded, Gender

Count Row %	a:	ai	
males	1730 95.69%	78 4.31%	1808
females	361 91.16%	35 8.84%	396
	2091	113	224

Significant at p = .0006

Also as in previous analyses, the factor group birthdate is examined more closely for clusters within the data. As in analysis of the data set excluding *I* and tokens containing the root *like*, speakers from all years produce the monophthong at statistically similar rates.

With respect to following phonological environment, exclusion of the tokens of *right* has no effect on the distribution of the monophthong among these factors except for the voiceless obstruents factor. Speakers produce the monophthong at 83.45 percent in the voiceless obstruent factor, a small change from 82.83 percent in the previous analysis of data excluding *I* and tokens containing the root *like*. Results are displayed below in Table 54.

Table 54: *I* and the Roots *Like, Right* Excluded, Following Phonological Environment

Count Row %	a:	ai	
voiced obstruents	492 97.81%	11 2.19%	503
nasals	650 97.01%	20 2.99%	670
word boundary	437 96.47%	16 3.53%	453
vowels	32 94.12%	2 5.88%	34
liquids	238 93.70%	16 6.30%	254
voiceless obstruents	242 83.45%	48 16.55%	290
	2091	113	2204

Significant at p less than .0001

Voiced obstruents and nasals exhibit the monophthong at a statistically similar rate and are recoded into one factor. The word boundary factor is also statistically similar to the voiced obstruents and nasals factors. In Bybee's (2007) volume of frequency studies, she notes that word boundary often behaves similarly to following obstruents because twice as many words begin with obstruents as vowels. Therefore, word boundary is combined with voiced obstruents and nasals. Vowels and liquids are also statistically similar to each other and thus are combined. The recoded data of following phonological environment is shown below in Table 55. Results suggest an implicational hierarchy of voiced obstruents, nasals, and word boundary > vowels, liquids > voiceless obstruents, very similar to Head's (2003) reported results.

Table 55: *I* and the Roots *Like, Right* Excluded, Following Phonological Environment, Recoded

Count Row %	a:	ai	
voiced obstruents, nasals, word boundary	1579 97.11%	47 2.89%	1626
vowels and liquids	270 93.75%	18 6.25%	288
voiceless obstruents	242 83.45%	48 16.55%	290
	2091	113	2204

Significant at p less than .0001

As before, the remaining internal factor groups, though not significant in the initial logistic regression, are examined individually to see if exclusion of tokens affects bivariate results. In this analysis, the four frequency quartiles exhibit decreasing percentages of the monophthong as frequency decreases, consistent with previous research (Bybee, 2007; Phillips, 1998, 2000). Results are displayed below in Table 56.

Table 56: *I* and the Roots *Like, Right* Excluded, Frequency

Count Row %	a:	ai	
greater than 10,000	56 98.25%	1 1.75%	57
5,000 – 9,999	512 97.90%	11 2.10%	523
1,200 – 4,999	563 93.83%	37 6.17%	600
0 – 1,199	960 93.75%	64 6.25%	1024
	2091	113	2204

Significant at p = .0004

The two higher frequency factors are very similar in their distribution, as are the two lower frequency factors. Pairwise testing confirms that the quartiles greater than 10,000 and 5,000 to 9,999 are statistically similar. The two lower quartiles are also statistically similar to each other. Thus, the four factors are recoded into two: greater than 5,000 and less than 5,000, shown in Table 57.

Table 57: *I* and the Roots *Like, Right* Excluded, Frequency, Recoded

Count Row %	a:	ai	
greater than 5,000	568 97.93%	12 2.07%	580
less than 5,000	1523 93.78%	101 6.22%	1624
	2091	113	2204

Significant at p less than .0001

With respect to word class, exclusion of tokens containing the root *right* dramatically increases the rate of production of the monophthong in the adverb factor, but otherwise the individual factors are nearly the same as in the analysis of the data set excluding only *I* and the root *like.* Results are shown below in Table 58.

Table 58: *I* and the Roots *Like, Right* Excluded, Word Class

Count Row %	a:	ai	
adverb	70 97.22%	2 2.78%	72
pronouns	501 96.91%	16 3.09%	517
prepositions	79 96.34%	3 3.66%	82
verbs	333 95.42%	16 4.58%	349
nouns	940 94.47%	3 3.66%	999
conjunctions	21 91.30%	2 8.70%	23
adjectives	147 88.55%	19 11.45%	166
	2264	151	2415

Significant at p = .0069

As in the analysis excluding *I* and tokens containing the root *like*, pronouns and prepositions, both function words, are statistically similar and are combined. In addition, verbs and nouns, both content words, are statistically similar and combined. The recoded factor group is shown in Table 59.

Table 59: *I* and the Roots *Like, Right* Excluded, Word Class, Recoded

Count Row %	a:	ai	
adverbs	70 97.22%	2 2.78%	72
pronouns and prepositions	580 96.835	19 3.17%	599
verbs and nouns	1273 94.72%	71 5.28%	1344
conjunctions	21 91.30%	2 8.70%	23
adjectives	147 88.55%	19 11.45%	166
	2264	151	2415

Significant at p = .0018

With respect to the factor group stress, there are very few unstressed tokens remaining in this data set, only two of which are diphthongs. Thus, this factor group is not included in the final logistic regression.

A final logistic regression reveals that only following phonological environment (p less than .0001), birthplace (p = .0002), and gender (p = .0007) significantly condition the alternation between the monophthong and the diphthong in this last data set. Recoding of the factor groups results in a reported R Square of 12 percent, a smaller percentage than in the initial logistic regression.

Summary of Results

Table 60: Significant Values, Final Logistic Regressions, All Analyses

	All Data (p value)	**All Data, No *I* (p value)**	**All Data, No *I* or Root Containing *Like* (p value)**	**All Data, No *I* or Roots Containing *Like, Right* (p value)**	***I* Only (p value)**	**Like Only (p value)**
N	5,121	2,637	2,415	2,204	2,484	222
Rate of Production of Monophthong	92.60%	89.38%	93.75%	94.87%	96.01%	41.89%
Birthplace	< 0.0001	0.0000	0.0009	0.0002	0.0140	0.0007
Following Phonological Environment	< 0.0001	0.0000	0.0000	< 0.0001	NA	NA
Word Class	0.0365	0.0007	NS	NS	NA	0.0027
Frequency	0.0023	0.0042	NS	NS	NA	NA
Birthdate	0.0004	0.0204	NS	NS	< 0.0001	NS
Gender	NS	NS	0.0333	0.0007	0.0231	NS
Stress	< 0.0001	NS	NS	NS	< 0.0001	NS
Initial R^2 Value	0.22	0.29	0.14	0.14	0.09	0.18
Final R^2 Value	0.22	0.27	0.14	0.12	0.10	0.17

*NS = Not significant
*NA = Not applicable

In summary, as shown in Table 60, all factor groups that are investigated in this study but gender are initially found significantly to condition the alternation between [a:] and [ai]. After recoding, 22 percent of the variation in this full data set is accounted for by the factor groups. Nearly half of the full data set, 48.51 percent, consists of tokens of *I*, so these tokens are analyzed separately. Analysis of the remaining tokens reveals that stress does not significantly condition the alternation between the monophthong and the diphthong; the remaining factor groups of gender, region, following phonological environment, word class, and frequency significantly correlate with variant production. Exclusion of the *I* tokens improves the R Square from 22 percent in the full data set to 27 percent in the final logistic regression of this smaller data set.

In this data set, the results for frequency are not consistent with previous research (Bybee, 2007; Phillips, 1998, 2000). The lowest frequency words exhibit a higher percentage of the monophthong than all but the highest frequency words. Investigation of the third quartile reveals that tokens containing the root *like* disfavor the monophthong, skewing this analysis. With removal of tokens containing *I* and the root *like* from the data set, gender, birthplace, and following phonological environment emerge as significant conditioning factor groups. However, the factor groups account for a smaller percentage of the variation, with R Square at 14 percent in the final logistic regression of the data set excluding tokens containing *I* and *like*.

After excluding these two types of tokens, frequency is not significant in multivariate analysis. Nevertheless, the lowest frequency words did have a significantly higher rate of production of the monophthong than the second quartile, prompting further investigation of the data set. The second quartile is found to contain a number of tokens containing the root *right* that do not favor the monophthong as highly as the other tokens not excluded. Therefore, tokens containing the root *right* are excluded from analysis. In a logistic regression of this reduced data set excluding tokens of *I* and containing the roots *like* and *right*, only gender, birthplace, and following phonological environment condition the alternation between the monophthong and the diphthong. These factor groups account for 12 percent of the variation in the speech of European-American textile mill workers employed in Columbus mills.

In the data set including only *I,* birthplace, gender, and stress condition the alternation between the monophthong and the diphthong. Factor groups analyzed in this data set account for 10 percent of the variation between the monophthong and the diphthong. In the data set of tokens containing the root *like*, birthplace and word class condition the alternation, with factor groups accounting for 17 percent of the variation. In the data set containing the root *right*, none of the factor groups significantly conditions the variation.

V. Conclusions, Implications, and Limitations

Conclusions

Overall, the speakers included in this study produce the monophthong at a very high rate. The analysis conducted after exclusion of the tokens *I* lowers this rate from 92.60 percent to 89.38 percent, but further analysis after excluding tokens containing the roots *like* and *right* confirms the initial finding. The final logistic regression, which identifies only the significant conditioning factor groups of following phonological environment (p less than 0.0001), birthplace ($p = 0.0002$), and gender ($p = 0.0007$), indicates that the speakers in this study produce the monophthong at a rate of 94.71 percent.

Although the R Squares indicate that the factor groups count for less than thirty percent of the variation, Wang (1969) points out that fewer factor groups condition variation as sound changes near completion. The high rate at which the speakers produce the monophthong coupled with the small number of significant conditioning factor groups suggest that monophthongal variant has been firmly established in this area of the Southeast for some time.

Stable sociolinguistic variation is characterized by males producing the nonstandard variant at a higher rate than females do (Labov, 1990). That is the case here; in all analyses, males produce the monophthong at higher rates than females do. This, coupled with the high frequency of the monophthongal variant, suggests that the monophthong is well established.

Wang's (1969) discussion of sound change points out that as a sound change progresses through the lexicon, the change becomes more regular, but that some lexical items lag behind the others. In this data set, tokens containing the roots *like* and *right* trail the remainder of the data. Tokens containing the root *like* exhibit the monophthong at a rate of only 41.89 percent, less than half the rate of production of the monophthong of the remaining data (94.87 percent) and slightly more than half the rate of production of the monophthong in the remaining voiceless obstruents (82.83 percent). Tokens containing the root *right* are not as anomalous; they exhibit the monophthong at 81.82 percent, a rate which is not statistically significant from other voiceless obstruents but still impacts frequency results. This finding, with only two items lagging behind the remaining data, also points to a well-established alternation.

Even in well-established variation, various internal and external factor groups affect production of the variable. Prior research in other Southern communities suggests that following phonological environment is the internal factor group with the strongest effect on production of the monophthong. That is the case. No following phonological environment disfavors the monophthong. The monophthongal variant is least likely to be found before voiceless obstruents, but this factor does not disfavor the monophthong. In the final analysis of the data, voiceless obstruents are the only factor to exhibit the monophthong less than 90 percent of the time, though this rate (82.83 percent) is higher than that reported in some other studies. Like Head (2003), who studied speakers from Coffee County, Alabama, just south of Pike County, this study finds the association between the monophthong and voiceless obstruents to be firmly established. Head (2003) also reports an implicational hierarchy similar to the one indicated by the present study: nasals, voiced obstruents, word boundary > vowels and liquids > voiceless obstruents. Finally, studies also suggest that younger speakers produce the monophthong more frequently before voiceless obstruents than older speakers do (Anderson, 2002; Thomas, 2001). However, this study finds no significant effect of age on the sociolinguistic variable (ai). This, too, points to stable variation.

Other research indicates that age conditions production of the monophthong, with older speakers producing the monophthong more frequently than younger speakers do (i.e., Bailey & Bernstein, 1989; Bowie, 2001; Edwards, 1997; Fridland, 2003). However, the findings of this study do not support this conclusion. Using data drawn from four speakers born between 1847 and 1976, Bailey (1997) hypothesizes a date in the 1870s as the origination of the variable investigated in this study. However, when Texas opened for settlement in the 1830s and 1840s, many who migrated there came from or through Alabama (Otto, 1989), which was already well established by then (Southerland & Brown, 1989). Settlers brought their speech patterns with them as they moved to new areas. The speaker born in 1897 in Bailey's (1997) study has a shortened offglide; the oldest speaker included in this study, born in 1896, exhibits a high rate of the monophthong.

The high rate of production among speakers in this study, the few significantly conditioning factor groups, and the fact that only two lexical items lag begin the remaining data indicate a well-established alternation among these speakers. Results of this study can add to the ongoing discussion regarding the origin of Southern American English, but further research is needed to clarify Bailey's (1997) hypothesis.

The speakers included in this study identified themselves as coming from one of four areas: the Houston County area, the Bullock County area, Pike County, and Columbus. Traditionally, Bullock County is a Black Belt County and Houston County a Wiregrass County. Sources differ regarding whether Pike County is a Black Belt or Wiregrass County ("Archives of Wiregrass History and Culture," n.d., "BlackBelt Action Commission, 2004). Dialectology studies have associated these four places with different dialect areas. For Carver (1987), the Chattahoochee River divides the Alabama and Georgia locations. For Pederson et al. (1986), Columbus, Bullock County, and northern Pike County are located in the Coastal Black Belt area; southern Pike County and Houston County are located in the Central Piney Woods area. In the final analysis of the data discussed in the present study, speakers from Bullock and Houston counties produce the monophthong statistically similarly to speakers from Columbus. This finding is inconsistent with Carver's (1987) analysis band is not entirely consistent with the LAGS findings. This study supports Schneider's (1998) observation that the Chattahoochee River is not a discrete boundary.

With respect to the remaining factor groups, although not statistically significant, the results are consistent with previous research. In the final analysis, frequency does not significantly condition the alternation between [ai] and [a:]. However, as one would expect given the overall high rate of the monophthongal variant, when considered in isolation, the more frequent a word is, the more likely the word is to be produced as a monophthong. Moreover, frequency analyses are crucial to the investigation of this study's corpus. In the initial analyses, words with lower frequencies are not associated with lower rates of production of the monophthong, inconsistent with research by Bybee (2007) and Phillips (1998, 2000). An examination of the frequency factors 1,200 to 4,999 and 5,000 to 9,999 reveals the anomalous behavior of tokens containing the roots *like* and *right.* Exclusion of these tokens results in consistency with previous research and ultimately leads to only two frequency factors: fewer than 5,000 and greater than 5,000. Bybee (2007) reports that coding frequency other than high and low does not improve description of her data. However, had the data in this study not been grouped in quartiles rather than halves, the impact of tokens containing the roots *like* and *right* might not have been discerned.

With respect to word class, initial analyses in this study are consistent with previous research (Bowie, 2001; Head, 2003) in that content words are more highly associated with the monophthong than are function words, except in the case of pronouns. Exclusion of tokens containing the roots *like* and *right* affects this finding, resulting in the following implication hierarchy: adverbs > pronouns and prepositions > verbs and nouns > conjunctions. This hierarchy is slightly different from results published in earlier research(Bowie, 2001; Eckert, 1996; Hazen, 2000; Head, 2003).

Stressed words are more closely associated with the monophthong than unstressed words are, consistent with both Bowie (2001) and Head (2003). That stress is not a significantly conditioning factor group after excluding the tokens of *I* suggests that the effects of stress are due primarily to this pronoun. Exclusion or limitation of this pronoun in other studies (Bowie, 2001; Eckert, 1996; Head, 2003) results in stress not being identified as a significant factor group.

Studies that discuss treatment of *I* in analysis either exclude the tokens altogether due to their categorical or near categorical behavior (Eckert, 1996; Head, 2003) or limit them due to their high frequency (Bowie, 2001). In this study, tokens of *I,* are pronounced as a monophthong at a rate of 96.01 percent. Nearly half the data set consists of these tokens. Analysis of the *I* data set indicates that birthplace, birthdate, gender, and stress significantly condition the alternation. Removal of these tokens improves the percentage of variation accounted for by the model (22 percent to 27 percent). However, subsequent removal of the tokens containing the roots *like* and *right* reduces this amount to 12 percent.

Implications and Limitations

The alternation between the monophthong and the diphthong is firmly established among the speakers included in this study and is in stable sociolinguistic variation. Nevertheless, the results and the limited nature of the subject pool due to the retrospective nature of this study point to areas where further research can be conducted.

Two of the external factor groups that other researchers have reported as significantly conditioning the sociolinguistic variable (ai) could not be investigated in this study. The participants were all European Americans employed at hourly wages at Columbus textile mills. Therefore, the effects of ethnicity and socioeconomic status could not be investigated. Further research that includes African-American speakers may reveal the relevance of ethnicity in this region of the Southeast. Additionally, including speakers of other socioeconomic statuses would allow investigation of the role of this factor on the alternation between the monophthong and the diphthong.

Dialectology studies and this study disagree on dialect boundaries in the areas of Alabama and Georgia from which speakers in this study originate. Further investigation into communities around the Chattahoochee River might clarify those dialect boundaries. In regard to urbanization, there is little support for Thomas's (2001) findings that urbanity influences production of the monophthong. Except for the speakers from Pike County, there is not a significant difference between speakers originally from farms and originally from Columbus. A study of speakers of different ages from these areas may reveal patterns not evident in this data set.

Because of the retrospective nature of the study, individual factors in the factor groups birthplace and gender are not evenly distributed. In particular, both the Bullock County speakers are female, and both the Pike County speakers are male. Only one speaker from Columbus is female. Including more speakers of both genders from each birthplace may reveal further patterns in the relationships among birthplace, gender, and the sociolinguistic variable (ai).

More industrial and more prosperous, the South has changed considerably since the language in this study was produced. Since the 1980s, Northerners have added considerably to the population across the South. Like the rest of the South, Columbus has grown. The location and relocation of a number of automotive industries to both Alabama and Georgia are contributing to the areas' expansion; the imminent military base consolidation in Columbus is expected to contribute further. Previous research has found that production of the monophthong across the South is changing (i.e., Bailey & Bernstein, 1989; Head, 2003; Fridland, 2003; McNair, 2005), with younger speakers producing the monophthong differently than older speakers do. This study does not find a significant effect of age on the variable (ai) and does find the correlation between voiceless obstruents and the monophthong firmly established among the speakers included in this study. However, the youngest speaker included in this study was born in 1935, before large-scale migration of Northerners south began. Further research should include speakers born since the Great Depression and the demographic changes in the South in order to investigate further the effect of age on production of (ai).

The data set on which this study is based contains only four tokens with following glides. The factor containing following vowels, while more numerous than tokens containing following glides, still has fewer tokens (N=34) than the next smallest factor, liquids (N=254). In order to investigate the impact of following phonological environment on production of the monophthong more fully, further research should attempt to garner more examples of these two factors. Asking participants to read word lists and paragraphs containing more examples of these factors would guarantee supplementary data for analysis.

In addition, this corpus contains a relatively high number of textile-specific terms containing (ai). Their assigned frequencies according to *The American Heritage Word Frequency Book* are relatively low. Interviewing speakers who are not so closely involved with the textile industry may affect the relative frequency of textile-specific terms in an expanded corpus. The effect of these tokens on the variable investigated in this study should be further investigated.

In this study, word frequency was coded in greater detail than in prior research. Without doing so, the anomalous behavior of tokens containing the roots *like* and *right* would likely have remained hidden. This, in conjunction with Neu's (1980) analysis, reminds us that it is important to investigate tokens skewing results, even if they are not categorical or nearly categorical in nature. Studies of the monophthongal (ai) in other communities should keep these results in mind to investigate whether the roots *like* and *right*, or others, affect production of the monophthong as they do in this study.

With respect to the factor group word class, exclusion of the tokens containing the roots *like* and *right* especially reduces the number of conjunctions, with only 23 of the original 122 remaining. Prepositions and adverbs are also affected. As is the case for following phonological environment, word lists and paragraphs would be helpful in garnering additional examples of conjunctions, prepositions, and adverbs not of the type excluded in this study in order to investigate the role of word class more fully in this alternation.

Though stress is not a significantly conditioning factor group once the tokens *I* were excluded, the factor group's strong association with the pronoun could be further investigated in other communities.

In the initial analysis, the model accounts for 22 percent of variation. As tokens are excluded, the amount of variation accounted for by the model decreases; by the final analysis, the reported R Square is only 12 percent. Including more data as suggested here may improve this finding.

Many people have rightly assumed that [a:] is a hallmark of Southern speech; the Southerners included in this study use the monophthong at a high rate, and very likely have done so for most of their lives. However, as this study has shown, even at the high rate of production of the monophthong exhibited by these speakers, the variation is conditioned by linguistic and social factor groups. Further research can only increase our understanding of the variable nature of this hallmark, highly salient feature of Southern speech.

References

Anderson, B. L. (1999). Source-language Transfer and Vowel Accommodation in the Patterning of Cherokee English /ai/ and /oi/. *American Speech*, *74* (4), 339 – 368.

Anderson, B. L. (2002). Dialect Leveling and /ai/ Monophthongization among African American Detroiters. *Journal of Sociolinguistics*, *6*(1), 86 – 98.

The Archives: Mill Workers Oral Histories. (no date). Retrieved August 25, 2006, from http://archives.colstate.edu/findingaids/mc109.shtml#a23

Archives of Wiregrass History and Culture. (no date). Retrieved June 22, 2009, from http://dothan.troy.edu/archives/Default.htm.

Bailey, G. (1997). When did Southern English Begin? In Edgar W. Schneider (ed.), *Englishes around the World, Vol. 1. Studies in Honor of Manfred Gorlach*. (pp. 255 – 275). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Bailey, G. and Bernstein, C. (1989). Methodology of a Phonological Survey of Texas. *Journal of English Linguistics*, *22*(1), 6 – 16.

Bailey, G. and Thomas, E. (1998). Some Aspects of AAVE Phonology. In *The Structure of African American Vernacular English* (pp. 85 – 109). In Mufwene, S., Rickford, J., Bailey, G., and Baugh, J. (eds.) *African-American English: Structure, History, and Use*. London: Routledge.

Bailey, G. and Tillery, J. (1996). The Persistence of Southern American English. *Journal of English Linguistics, 24*(4), 308 – 321.

Barnes, B. (2008, May 25). "Diamond in the Rough." *Columbus Ledger-Enquirer,* Explore p. 6.

Bernstein, C. (1993). Measuring Social Causes of Phonological Variation in Texas. *American Speech* 68, 227 – 240.

Bibb Manufacturing Company. (2006, October 13). In *The New Georgia Encyclopedia.* Retrieved November 13, 2006, from http://georgiaencylopedia.org/nge/Article.jsp?id=h-3213.

BlackBelt Action Commission. (2004, August 14). Retrieved June 22, 2009 from http://blackbeltaction.alabama.gov/EO.aspx.

Bowie, D. (2001). The Diphthongization of /ay/: Abandoning a Southern Norm in Southern Maryland. *Journal of English Linguistics, 29* (4), 329 – 345.

Bybee, J. (2000). Phonological Evidence for Exemplar Storage of Multiword Sequences. *Second Language Acquisition*, 24, 215 – 221.

Bybee, J. (2007). *Frequency of Use and the Organization of Language.* New York: Oxford University Press.

Carroll, J. B., Davies, P., and Richman, B. (eds.) (1971). *The American Heritage Word Frequency Book.* Boston: Houghton Mifflin.

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

Cedergreen, H. and Sankoff, D. (1974). Variable Rules: Performance as a Statistical Reflection of Competence. *Language* 50, 333 – 355.

Center for Business and Economic Research. (no date).“Traditional Counties of the Alabama Black Belt.” Tuscaloosa, AL: University of Alabama.

Chambers, J. (2002). *Sociolinguistic Theory: Second Edition.* Malden, MA: Blackwell Publishing.

Chi Square, Cramer’s V, and Lambda for a Rows by Contingency Table. (2009). *Vassar Stats*. Retrieved 2007 – 2009 from http://faculty.vassar.edu/lowry/newcs.html.

The City of Dothan, Alabama. (2006). History. Retrieved June 22, 2009 from http://www3.dothan.org/history/index.html.

Columbus. (2004, March 30). In *The New Georgia Encyclopedia.* Retrieved November 13, 2006 from http://georgiaencyclopedia.org/nge/Article.jsp?id=h-2208.

Creek Indians. (2002, February 2). In *The New Georgia Encyclopedia.* Retrieved June 22, 2009 from http://www.georgiaencyclopedia.org/nge/Article.jsp?id=h-579&hl=y.

Crane, L. B. (1977). The Social Stratification of /ai/ in Tuscaloosa, Alabama. In David L. Shores and Carole P. Hines (eds.), *Sociolinguistic patterns in British English* (pp. 37 – 51). London: Edward Arnold.

DeVinne, P. B. (ed.) (1982). *The American Heritage Dictionary: Second College Edition.* Boston: Houghton Mifflin Company.

Eckert, P. (1996). (ay) Goes to the City: Exploring the Expressive Use of Variation. In G. Gregory, C. Feagin, D. Schiffirin, & J. Baugh (eds.), *Variation and change in language and society.* Philadelphia: John Benjamins Publishing Company.

Edwards, W. F. (1997). The Variable Persistence of Southern Vernacular Sounds in the Speech of Inner-city Black Detroiters. In C. Bernstein, T. Nunnally, and R. Sabino (eds.), *Language variety in the South revisited,* (pp. 76 – 86). Tuscaloosa, AL: The University of Alabama Press.

Feagin, C. (1997). The African Contribution to Southern States English. In C. Bernstein, T. Nunnally, & R. Sabino (eds.), *Language Variety in the South Revisited* (pp. 123 – 139). Tuscaloosa, AL: The University of Alabama Press.

Feagin, C. (2003). Vowel Shifting in the Southern United States. In S. J. Nagle and S. L. Sanders (eds.) *English in the Southern United States* (pp. 126 – 140). Cambridge, UK: Cambridge University Press.

Fisher Exact. (n.d.) Retrieved 2007 – 2009 from http://www.quantitativeskills.com/sisa/statistics/fisher.htm.

Fridland, V. (2000). The Southern Shift in Memphis, Tennessee. *Language Variation and Change*, *11*, 267 – 285.

Fridland, V. (2003). 'Tie, tied, and tight': The Expansion of /ai/ Monophthongization in African-American and European-American Speech in Memphis, Tennessee. *Journal of Sociolinguistics*, *7*(3), 279 – 298.

Green, M. D. (1982). *The Politics of Indian Removal: Creek Government and Society in Crisis.* Lincoln, Nebraska: University of Nebraska Press.

Gregory, M., Raymond, W. D., Bell, A., Foster-Lussier, E., and Jurafsky, D. (1999). The Effects of Collocational Strength and Contextual Predictability in Lexical

Production. In Bybee, J. *Frequency of Use and the Organization of Language.* New York: Oxford University Press.

Hall, J. D., Leloudis, J., Korstad, R., Murphy, M., Jones, L., and Daly, C. B. (1987). *Like a Family: The Making of a Southern Cotton Mill World.* Chapel Hill, NC: University of North Carolina Press.

Hazen, K. (2000). A Methodological Suggestion on /aj/ Ungliding. *American Speech, 75*(2), 221 – 224.

Head, A. (2003). *The Monophthongization of /ai/ in Elba and the Environs.* Unpublished master's thesis, Auburn University, Alabama.

Kurath, H. (1949). *A Word Geography of the Eastern United States.* Ann Arbor: University of Michigan Press.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

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. and Ash, S. (1997). Understanding Birmingham. In C. Bernstein, T. Nunnally, & R. Sabino (eds.), *Language Variety in the South Revisited* (pp. 508 – 573). Tuscaloosa, AL: The University of Alabama Press.

Lance, D. M. (1994). Variation in American English. In Donald M. Lance and Stewart A. Kingsbury, *American Pronunciation*, 12th ed. (pp. 345 – 373). Ann Arbor, Michigan: University of Michigan Press.

McMillan, J. and Montgomery, M. (1989). *Annotated Bibliography of Southern American English*. Tuscaloosa, AL: The University of Alabama Press.

McNair, E. (2005). *PADS 90: Mill Villagers and Farmers: Dialect and Economics in a Small Southern Town.* Durham, NC: Duke University Press.

Mendoza-Denton, N., Hay, J., and Jannedy, S. (2003). Chapter 4: Probabilistic Sociolinguistics: Beyond Variable Rules. In R. Bod, J. Hay, and S. Jannedy *Probabilistic Linguistics* (pp. 97 – 138). Cambridge, MS: MIT Press.

Mufwene, S. (2003). The Shared Ancestry of Southern Englishes. In S. J. Nagle & S. L. Sanders, *English in the Southern United States* (pp. 64 – 81). Cambridge, UK: Cambridge University Press.

Neu, H. (1980). Ranking of Constraints on /t,d/ deletion in American English: A Statistical Analysis. In W. Labov (ed.), *Locating Language in Time and Space* (pp. 37 – 54). New York: Academic Press.

Otto, J. S. (1989). *The Southern Frontiers, 1607 – 1860: The Agricultural Evolution of the Colonial and Antebellum South.* New York: Greenwood Press.

Pederson, L., McDaniel, S. L., Bailey, G. and Bassett, M. (1986). *Linguistic Atlas of the United States.* Athens, GA: University of Georgia Press.

Phillips, B.S. (1994). Southern English Glide Deletion Revisited. *American Speech*, 69(2), 115 – 127.

Phillips, B. S. (1998). Lexical Diffusion is not Lexical Analogy. *Word*, 49(3), 369 – 381.

Phillips, B. S. (2000). Fast Words, Slow Words. *American Speech*, 75(4), 414 – 416.

Rand, D. and Sankoff, D. (1990). Goldvarb Version 2: A Variable Rule Application for MacIntosh. http://albbuquerque.bioinformatics.uottawa.ca/GoldVarb/GoldManual.dir/index.html.

Sabino, R. Frequency, Language Learning, and Language Change. A paper presented at SECOL 2007.

Sabino, R., Diamond, M. S., and Oggs, A. H. (2004). A Quantitative Study of Plural Marking in Three Non-Urban African American Language Varieties. A paper presented at LAVIS III Tuscaloosa, AL.

Schilling-Estes, N. (2000). Investigating Intra-ethnic Differentiation; /ai/ in Lumbee Native American English. *Language Variation and Change, 12*, 141 – 174.

Schneider, E. W. (1998). The Chattahoochee River: A Linguistic Boundary? In M. B. Montgomery and T. E. Nunnally (eds.) *From the Gulf States and Beyond: The Legacy of Lee Pederson and LAGS* (pp. 123 – 146). Tuscaloosa, AL: The University of Alabama Press.

Schneider, E. W. (2003). Shakespeare in the Coves and Hollows? Towards a History of Southern English. In S. J. Nagle and S. L. Sanders (eds.) *English in the Southern United States* (pp. 17 – 35). Cambridge, UK: Cambridge University Press.

Selig Center for Economic Growth. (no date.) “County Map of Georgia.” Athens, GA: University of Georgia.

Southerland, Jr., H. D. and Brown, J. E. (1989) *The Federal Road through Georgia, the Creek Nation, and Alabama, 1806 – 1836.* Tuscaloosa, Alabama: The University of Alabama Press.

Tagliomonte, S. A. (2006). Analyzing Sociolinguistic Variation. Cambridge: Cambridge University Press.

Tillery, J. and Bailey, G. (2003). Urbanization and the Evolution of Southern American English. In S. J. Nagle and S. L. Sanders (eds.) *English in the Southern United States* (pp. 59 – 172). Cambridge, UK: Cambridge University Press.

Timeline: Texas Settlement History. (2004). Retrieved June 22, 2009 from http://www.pbs.org/wgbh/amex/alamo/timeline/index.html.

Thomas, E. (1997). A Rural/Metropolitan Split in the Speech of Texas Anglos. *Language Variation and Change,* 9(3), 309 – 332.

Thomas, E. (2001). *PADS 85: An Acoustic Analysis of Vowel Variation in New World English.* Duke, NC: Duke University Press.

Thomas, E. and Bailey, G. (1998). Parallels between Vowel Subsystems of African American Vernacular English and Caribbean Anglophone Creoles. *Journal of Pidgin and Creole Languages,* 13, 267 – 296.

Tillery, J. and Bailey, G. (2003). Approaches to Real Time in Dialectology and Sociolinguistics. *World Englishes,* 22, 351 – 365.

Union Springs. (2006). Our History. Retrieved June 22, 2009 from http://www.unionspringsalabama.com/history.html.

Wang, W. S-Y. (1969). Competing Changes as a Cause of Residue. *Language*, 45(1), 9 – 25.