Chapter 3

Data Collection

3.1 Introduction

In this chapter I will describe the procedures I employed to collect the data for this dissertation. First, Section 3.2 will discuss the methodology in recruiting speakers to participate in interviews. Due to this project's non-traditional method of vowel analysis (involving transcription of the interviews and forced alignment) as well as the demographic characteristics of the desired speakers, the methods I employed for contacting speakers were different than the methods used for many other sociolinguistic studies. Next, Sections 3.3 and 3.4 will describe the types of speakers I targeted for inclusion in the corpus, and provide tables with demographic information for every speaker whose data was analyzed. Then, Section 3.5 will briefly summarize the demographic characteristics of the corpus as a whole. Section 3.6 will describe the interview procedure and the formal methods used for targeting specific variables. Finally, Sections 3.7 and 3.8 will provide a high-level description of the methodologies for data analysis involving transcription, forced alignment, and automatic vowel analysis (for a more technical description of this methodology, see Chapter 4).

3.2 Selection of Individuals

While great care is often taken in the selection of individual speakers for inclusion in a sociolinguistic corpus to ensure a random, stratified sample including the social categories of interest (e.g., (Sankoff and Sankoff 1973)), this is not usually possible to do in a dialect geography study. Due to the increased time and effort involved in sampling speakers from a larger geographical area, only a handful of speakers are usually selected in each community. Thus, it is not possible to adequately control for speaker characteristics such as age, sex, and socioeconomic status. Rather, speakers in such studies are usually selected by a somewhat ad hoc process, as described in DARE (Cassidy and Hall 1985:xiv): "the intention was to maximize the collection of materials by going to the places and people most likely to furnish the largest amount of appropriate data." This general approach was also followed in the sampling of speakers for this dissertation. In Erie, the focus was on older speakers, in an attempt to document the earliest stages of the city's shift away from the North. In other areas, the closest attention was paid to boundary regions, especially the town of Ripley, NY, in order to obtain a detailed apparent time view of how the boundaries have been changing recently.

One relatively easy way of conducting targeted sampling based on geographic location is to select individuals from phone book listings. This is the is the method of participant recruitment that was employed by the ANAE (Labov et al. 2006:24–27). The main benefit of this procedure is that it enables the researcher to obtain an extremely wide geographical coverage without needing to travel to conduct the research. However, there is one main drawback to this procedure: since the calls are unsolicited and unexpected, the researcher is not able to prepare the speaker in advance with any specific materials. Thus, it is impossible to use formal methods that require the speaker to have access to a physical object, such as word lists, reading passages, or picture naming materials. The ANAE overcame this

difficulty with a combination of verbal tasks, such as targeted elicitation of lexical items, naming of lexical items from a restricted set (e.g., numbers, days of the week, articles of clothing), semantic differentials, etc.—see Labov et al. (2006:32–34) for a complete list of all such formal elicitation tasks employed by the ANAE researchers. However, the best way to ensure that there is a large set of words that are uttered by all speakers in a corpus is to use a word list. The ANAE researchers realized this, and attempted to arrange follow-up interviews with speakers after sending them a word list in the mail (Labov et al. 2006:29); however, this second interview was only carried out with a minority of the ANAE speakers. Thus, in order to ensure that all speakers would participate in the word list and reading passage tasks, and to greatly speed up the minimal pairs task, I decided to not conduct any unsolicited telephone interviews for this dissertation. I did conduct several interviews over the telephone, as described in Section 3.6.2; however, in all cases the speakers had been contacted prior to the interview, and a specific appointment had been arranged (this was necessary to ensure that the speakers would have internet access at the time of the interview).

A second method of efficiently obtaining data for a dialect geography study is the Short Sociolinguistic Encounter (SSE), described in Ash (2002). This method involves anonymous face-to-face interviews of people in public places in the town of interest, thus enabling the use of word lists, reading passages, or any other printed formal method, during the initial contact. An SSE is usually shorter in duration than a full interview, thus enabling the researcher to collect data from several individuals during a single day of field work in a given town. To find subjects, the researcher approaches individuals in public places such as parks, cafes, stores, etc., and conducts the SSE in the same location. The technique was originally envisioned "to provide the maximum amount of data on a specified, small set of variables" (Ash 2002:2). For example, it was used successfully by Ash (2002) to obtain large amounts of data about the distribution of short-a in the Mid-Atlantic region,

and by Johnson (2007) to investigate the distribution of /ah/, /o/, and /oh/ along the border between Massachusetts and Rhode Island. Recent work by Dinkin (2009) has maintained the basic approach of the SSE, but extended it somewhat in order to obtain enough data to describe each speaker's entire vowel system.

While the SSE is an efficient method that enables the dialect geographer to obtain a decent sample of the speech in a given location in as little as a single day, it does have some drawbacks. First of all, since the interviews are usually conducted anonymously, the speakers can not be contacted again in the future if follow-up research is necessary. More importantly, the anonymity of the encounter makes it much more difficult to ask the informant for referrals to other potential participants. Such personal introductions are often crucial in enabling the researcher to come in contact with speakers of a targeted demographic group that might not be easy to encounter in an SSE. This point is important for this dissertation, since the employment of the SSE as the sole method of data collection would have made finding a large number of elderly speakers very difficult.

However, the most severe problem of the SSE is that the recordings are often of lower quality. Due to the nature of how speakers are met in an SSE, recordings often take place outside (e.g., in public parks), or in public establishments such as cafes or retail stores. In both types of environments, there is almost always a substantial level of background noise that is also picked up during the recording. For the purposes of a standard soiciophonetic analysis of vowel quality, this would not be a problem, since the manual extraction of vowel formants is not hindered greatly by the presence of such background noise. However, for the purposes of this dissertation, the technique of forced alignment was employed to automatically segment the audio signal into phonemes (see Section 4.2 for a description of this methodology) which, in turn, enabled the automatic extraction of vowel formants. For the forced alignment procedure to perform optimally, the speech signal should be as clean as possible; background noise can be mis-recognized as speech from the speaker,

thus causing errors in phoneme segmentation. For this reason, I attempted to conduct interviews in locations with a minimal amount of background noise, ideally in the speaker's residence, or a quiet public place, such as a meeting room in a library. Such interviews are not normally possible when using the SSE methodology.

In order to overcome the drawbacks of random telephone interviews and the SSE, a somewhat more laborious method was chosen as the primary way of selecting speakers. Advertisements were put up in libraries and community centers in all of the towns of interest, and potential informants were asked to contact me by telephone or email if they were interested in participating. Upon initial contact, a bried demogrphic interview was conducted to make sure that the speaker was born in the town of interest and lived there continuously until the age of 18. An effort was made to screen for speakers who also lived their entire adult lives in the town of interest, but speakers were not excluded if they lived outside of the town for short periods in their adult lives (for example, during college or for a job transfer). If the speaker met the residency criteria, a time would be arranged for the interview, either at the speaker's residence or at a maximally quiet public place. Response to the advertisements was surprisingly heavy, especially in the smaller towns, and many participants were initially contacted in this manner.

After conducting each interview, the participant was then asked to recommend any friends or family members who met the residency criteria and would be willing to participate. Almost all informants were willing to provide contact information for at least one other participant; when these references were pursued, they also usually led to more successful interviews. Thus, the bulk of the speakers were initially found through recommendations from previously interviewed speakers; usually only a single response to the initial advertisement in a given town was required to obtain an endless string of potential speakers. The only interviews that were not conducted in this manner were the ones at the Sun Valley retirement community (described below in Section 3.3.1), a few initial in-

terviews in the city of Erie that were obtained through personal contacts, and a few early SSE's in neighboring towns that were conducted before the negative effects of the poorer quality recordings were fully realized. Despite the fact that the chosen methodology required a larger investment of time in speaker recruitment, and meant that more time had to be spent in the field, I believe the resulting increased quality of the recordings and deeper personal connection with the participants made the effort worthwhile.

Finally, in the later stages of the field work, several interviews were also conducted over the telephone using a call collection interface provided by the Linguistic Data Consortium. Speakers were initially contacted in the same manner as described above, but were interviewed over the telephone while I was in Philadelphia. This was done simply to reduce the number of trips that I would have to take to the area. The process of conducting telephone interviews through the LDC's interface led to very high quality recordings for the purpose of vowel analysis, despite the restricted frequency range transmitted over the telephone. Background noise in these interviews is almost non-existent, and the LDC's interface separates the input from the two phone lines into two separate audio files; both of these factors led to easier transcription and improved performance of the forced alignment system.

3.3 Selection of Speakers: Time Depth

The first goal of selecting speakers for analysis was to push the time depth for our knowledge of Erie as far back as possible. As shown in Chapter 6, the data from the two LAMSAS speakers suggests that the merger of /o/ and /oh/ occurred in Erie sometime after 1910. In order to test this hypothesis, an effort was made to record elderly Erieites with the hope of finding some who were born before the merger took place. If present-day speakers with the /o/ $\sim$ /oh/ distinction could be found, it would enable us to pinpoint the date of the merger with a high degree of certainty, and, thus, convincingly demonstrate when Erie ceased to be

a Northern city. In addition, data was obtained from several archival sources in an attempt to collect data from an earlier time period than is possible with living speakers. This real time evidence complements the apparent time evidence obtained from my own fieldwork and provides a more complete picture of the course of linguistic change in Erie and the neighboring region.

3.3.1 Sun Valley residents

As discussed above, my first goal in data collection was to find several elderly Erieites who were born before or around the date suggested by the LAMSAS evidence for the merger of /o/ and /oh/ in Erie. The set of data providing this apparent time evidence comes from one-on-one interviews I conducted in person at an upscale retirement community in Erie, which I will call Sun Valley. I contacted the facility's commuity director, and she put an announcement in the weekly newsletter advertising my survey. Volunteers who were interested in participating in the interviews contacted her, and she set up appointments throughout the course of October, 2007. In total, I conducted 12 interviews at Sun Valley with life-long residents of Erie and three other cities of interest; these speakers ranged in age from 66 to 95. The oldest Sun Valley resident was born in 1912, and would thus be a good candidate for maintaining the distinction, assuming the chronology in Section 2.3 is correct. Table 3.1 displays the demographic characteristics of these 12 speakers.

3.3.2 Archival Material

In addition to interviewing the older Erie residents from Sun Valley, I made an attempt to obtain real time data from speakers that would extend the time depth even further into the past. For this purpose, archival material was obtained from three distinct sources described below.

Name	Born	City	State	Occupation
Dan R.	1912	Erie	PA	engineer
Robert E.	1916	Erie	PA	doctor
Mary D.	1919	Erie	PA	
Flora R.	1919	Erie	PA	teacher
Eloise B.	1925	Erie	PA	
Charles B.	1925	Erie	PA	manager
Dottie A.	1926	Erie	PA	teacher
Sally W.	1928	Erie	PA	deputy sheriff
Dana W.	1941	Erie	PA	teacher
Jane S.	1915	Oil City	PA	
Marge K.	1919	Pittsburgh	PA	
Walter K.	1927	Buffalo	NY	engineer

Table 3.1: Demographic characteristics of the 12 Sun Valley residents

The first source was the Seasonal Workers in Viticulture (SWV) corpus. This corpus was compiled as part of an oral history project conducted in 1988 to document the local grape growing industry around North East, PA. The town of North East is located in the northeastern corner of Erie County, about 20 miles from downtown Erie, and directly across the state line from New York. The town has always been a center for grape production, with a focus on producing juice from the Concord variety. A few other grape varieties are also grown, and several wineries exist along both sides of Route 20. Viticulture has always been one main source of jobs for North East residents, in addition to other types of agriculture.

The SWV project attempted to interview older native residents of North East who had owned vineyards or who had worked as grape pickers, although a few younger people and a few in-migrants were also interviewed. In all, 50 recordings were made, each about one hour in length. The interviews are available to the public as cassette tapes at the Erie County Historical Society. I selected the two oldest, native North East residents from the corpus for analysis, since they were most likely to have maintained a distinction between /o/ and

ID	Born	Interviewed	City	State	Source
SWV 039	1906	1988	North East	PA	SWV
SWV 046	1907	1988	North East	PA	SWV

Table 3.2: Demographic characteristis of the two speakers from the SWV corpus whose speech was analyzed manually

/oh/. These two speakers were born in 1906 and 1907. Their demographic information is summarized in Table 3.2.

The recordings for these two speakers are quite poor: the microphone was positioned far away from the speaker's mouth, and each recording contains a large amount of background noise. They are thus poor candidates for the procedure involving forced alignment and automatic vowel analysis that will be used for the rest of my data. Therefore, the vowels for these two speakers were analyzed manually, and they will not be included in the Natural Break maps in Chapter 5. However, the manual measurements for these two speakers will be included in the analysis of the merger of /o/ and /oh/ in Erie in Chapter 6, since they provide a useful early source of data.

The second source of archival material is a set of recordings from the Dictionary of American Regional English (Cassidy and Hall 1985–2002). I analyzed the interview speech from three DARE speakers from Erie County (two from North East and one from Union City). In addition, I analyzed the "Arthur the Rat" reading passage recordings from these three speakers and an additional 11 DARE speakers from the boundary regions around Erie.[1] The demographic information for these 14 speakers is provided in Table 3.3. Personal names of the speakers in the DARE corpus are not public information; the names provided in Table 3.3 are pseudonyms that I created. The DARE ID numbers are also

[1]The digitized DARE audio files were provided by Joan Hall, chief editor of DARE, at the University of Wisconsin.

Name	ID	Born	City	State	Occupation
Nancy S.	PA129	1908	North East	PA	teacher
Sarah N.	PA130	1897	North East	PA	homemaker
Gladys T.	PA131	1899	Meadville	PA	policeman
Bill C.	PA133	1950	Meadville	PA	librarian
Agatha S.	PA181	1907	Warren	PA	teacher
Steven G.	PA182	1915	Warren	PA	medicine
Maggie S.	PA234	1900	Union City	PA	teacher
Anne B.	NY099	1898	Fredonia	NY	homemaker
Leslie B.	NY100	1897	Fredonia	NY	teacher
Wallace L.	NY101	1892	Fredonia	NY	factory worker
Jonas H.	NY102	1898	Ripley	NY	vintner
Clarence T.	NY103	1886	Ripley	NY	engineer
Jill C.	NY104	1889	Ripley	NY	seamstress
Ted L.	NY215	1904	Jamestown	NY	craftsman

Table 3.3: Demographic characteristics of 14 DARE speakers from archival sources whose acoustic data were analyzed, interviewed 1968 - 1969

Name	Born	City	State	occupation
H.O. Hirt	1887	Erie	PA	CEO

Table 3.4: Demographic characteristics of H.O. Hirt, interviewed in 1977 for the Erie Insurance Company archives

provided, to enable reference to the specific speakers in published DARE material.

Additionally, I obtained a VHS tape of an interview with H.O. Hirt, the founder of Erie Insurance Exchange.[2] Hirt was born in Erie in 1887, founded the company in 1925, and served as its CEO until 1976. The interview was conducted with him in 1977 and a 20-minute segment of it was released as a publicity tape by the Erie Insurance Group. This interview thus represents the oldest recorded Erieite that I have so far been able to discover. H.O. Hirt's demographic information is summarized in Table 3.4.

Finally, seven speakers from the ANAE corpus were reanalyzed for this dissertation

[2]The tape was given to me by the staff archivist at the Erie Insurance Group.

Name	ID	Born	City	State	Occupation
Irvin H.	TS168	1932	Cleveland	OH	special education
Samuela S.	TS364	1964	Erie	PA	car wash attendant
Ken K.	TS545	1961	Pittsburgh	PA	student
Gwen S.	TS355	1929	Pittsburgh	PA	unknown
Cecilia S.	TS356	1933	Pittsburgh	PA	student
Henry K.	TS544	1935	Pittsburgh	PA	teacher
Charlotte S.	TS739	1961	Pittsburgh	PA	secretary

Table 3.5: Demographic characteristics of the 7 ANAE speakers whose acoustic data were re-analyzed, interviewed 1994 - 1996

using the methodology of transcription and forced alignment. Most of the ANAE speakers chosen for reanalysis (five out of the seven) were from Pittsburgh. They were selected in order to provide a more complete description of this city for comparison with Erie, since my own fieldwork recordings only contained three speakers from Pittsburgh. Because the method of analysis for these seven ANAE speakers involved transcription and forced alignment, their vowel analyses are based on their entire interviews, not only the words that were measured manually by the ANAE annotators. Their demographic information is provided in Table 3.5. Again, personal names of the speakers in the ANAE corpus are private; the pseudonyms in Table 3.5 are the ones provided by the public version of the ANAE database released with the corpus.

3.4 Selection of Speakers: Geographical Depth

The second aim of selecting speakers for analysis was to collect data from the small towns around Erie, in an attempt to determine the nature of the dialect boundaries between Erie and the North, on the one hand, and Erie and Pittsburgh, on the other.

First of all, a more complete description of the city of Erie itself was needed, since the

Name	Born	Occupation
Barry G.	1938	car salesman
Laurie G.	1946	nurse
Pam R.	1945	homemaker
Sophie D.	1950	computer programmer
Jane L.	1953	gardener
Tom L.	1953	welder
Greg A.	1980	unemployed
Sally L.	1982	office worker

Table 3.6: Demographic characteristics of 8 speakers from the city of Erie

only study to date of Erie speech involving acoustic analysis is the ANAE, in which two speakers were analyzed. I recorded and analyzed the speech of nine native Erieites. Their demographic information is presented in Table 3.6.[3]

In order to investigate the boundary region between Erie and the North, I visited several towns in Chautauqua Co., NY, located in the western part of the stated along the border with Erie Co., PA. The speakers from these towns are listed in Table 3.7. As can be seen from Table 3.7, a strong focus was placed on the town of Ripley, NY. This town, the first one in NY after crossing the state line from PA, was discovered to have an apparent time distribution of /o/ and /oh/ indicating that these two phonemes began to merge about two generations ago. Thus, I sought a higher number of speakers there than in other towns in Chautauqua Co. in order to provide a more complete apparent time distribution of this change.

Due to the time spent on the field work in the boundary area between Erie and the North in Chautauqua Co., NY, I was only able to interview a few speakers from the other boundary area between Erie and the North, namely to the west of Erie in Ohio. These

[3]All names given for speakers that I interviewed are pseudonyms. They are provided to facilitate reference to the vowel plots for specific speakers that will be presented in Chapter 6 and 7.

Name	Born	Town	Occupation
Winifred S.	1925	Ashville	waitress
Mae S.	1925	Bemus Point	secretary
Bill R.	1930	Dunkirk	insurance salesman
Daisy T.	1921	Fredonia	secretary
Joan P.	1938	Jamestown	social worker
Barbara C.	1952	Jamestown	librarian
Amy G.	1963	Jamestown	teacher
Amy C.	1937	Westfield	librarian
Ralph O.	1934	Ripley	grape farmer
Margaret B.	1940	Westfield	historian
Stan R.	1948	Ripley	grape farmer
Rachel A.	1951	Ripley	daycare provider
Larry K.	1952	Ripley	town supervisor
John M.	1953	Ripley	town supervisor
Pam O.	1958	Ripley	winery owner
Daphne R.	1958	Ripley	grape farmer
Shelly I.	1960	Westfield	
Rachel C.	1963	Ripley	town clerk
Troy R.	1989	Ripley	student
Ryan N.	1994	Ripley	student
Grace N.	1997	Ripley	student
Jeff H.	1952	Buffalo	teacher

Table 3.7: Demographic characteristics of 23 speakers from New York (all speakers are from Chautauqua County except Jeff H.)

Name	Born	Town	County	Occupation
Patti N.	1957	Conneaut	Ashtabula	librarian
Brenda W.	1937	Ashtabula	Ashtabula	
Lisa C.	1940	Cleveland	Cuyahoga	librarian

Table 3.8: Demographic characteristics of 3 speakers from northeastern Ohio

Name	Born	Town	County	Occupation
Bob O.	1947	Ford City	Armstrong	teacher
Mary N.	1948	Butler	Butler	librarian
Abe M.	1944	Franklin	Venango	teacher
Ed W.	1932	Franklin	Venango	engineer
Bart P.	1946	Franklin	Venango	goverment administrator
Gary S.	1930	Franklin	Venango	goverment administrator
Carol H.	1942	Warren	Warren	health care administrator
Charlene O.	1934	Pymatuning	Crawford	teacher
Allison N.	1932	Pittsburgh	Allegheny	
Sara B.	1958	Pittsburgh	Allegheny	health care administrator
Kevin W.	1974	Greensburg	Westmoreland	banking

Table 3.9: Demographic characteristics of 11 speakers from western PA

speakers are listed in Table 3.8.

Additionally, research was conducted in the region of western PA between Erie and Pittsburgh in an attempt to ascertain the extent of the influence of the Pittsburgh system in the region. These speakers are listed in Table 3.9.

Finally, speakers were sought from most of the small towns in Erie Co., PA, especially those in the eastern and southern portions of the county, in an attempt to provide a more complete picture of the boundary areas between Erie and the North and Pittsburgh, respectively. These speakers are listed in Table 3.10.

Furthermore, I conducted abbreviated interviews with an additional 18 speakers who could not be recorded due to time constraints. Most of these speakers completed the por-

Name	Born	Town	Occupation
Jane W.	1948	Edinboro	principal
James N.	2000	Edinboro	student
Irene C.	1927	Wattsburg	telephone operator
Cathy A.	1955	Lawrence Park	librarian
Catherine F.	1942	North East	banker
Betty W.	1936	North East	teacher
Sharon N.	1931	Union City	
Cindy M.	1943	Girard	teacher
Charlotte S.	1955	Waterford	teacher
Marjorie S.	1986	Waterford	student

Table 3.10: Demographic characteristics of 10 speakers from Erie Co.

tions of the survey involving minimal pairs and acceptability judgments (see Section 3.6.1 for more information about the components of the interview), although a few only completed one or the other. Since there is no acoustic data available for these speakers, they will only be included in the maps for minimal pairs and lexical and morphosyntactic variables, when appropriate. Table 3.11 provides demographic information for these 18 speakers.

3.5 Characteristics of the Corpus

One of the main research goals of this dissertation is to explore the oldest stages of the Erie system, in order to discover when and how Erie ceased to be part of the North. Many of the speakers thus fit the profile of the NORM (Non-mobile, Older, Rural, Male) speaker that is traditionally the target speaker for a dialect geography study (Chambers and Trudgill 1999:29) (although many elderly female speakers were also interviewed). Younger speakers were only targeted specifically when a change in progress was detected in the community, and a more complete apparent time distribution was desired. Figure 3.1 displays a histogram of the birth year for all 106 speakers in the corpus, both from interviews con-

Name	Born	Town	County	State	Occupation
Susan B.	1959	Westfield	Chautauqua	NY	unknown
Sheila T.	1950	Ripley	Chautauqua	NY	waitress
Jane L.	1960	Ripley	Chautauqua	NY	waitress
Tracy N.	1972	Ripley	Chautauqua	NY	waitress
Heather E.	1990	Ripley	Chautauqua	NY	high school student
Trevor J.	1990	Ripley	Chautauqua	NY	high school student
Teri F.	1990	Ripley	Chautauqua	NY	high school student
Carrie B.	1990	Ripley	Chautauqua	NY	high school student
Vanessa T.	1990	Ripley	Chautauqua	NY	high school student
Chloe S.	1990	Ripley	Chautauqua	NY	high school student
Adam R.	1990	Ripley	Chautauqua	NY	high school student
Rebecca R.	1980	Ripley	Chautauqua	NY	baker
Charles S.	1956	Buffalo	Erie	NY	public services
Edith N.	1933	Ashtabula	Ashtabula	OH	clerk
Dan A.	1923	Erie	Erie	PA	retail manager
Tess E.	1945	Erie	Erie	PA	telephone operator
Laura S.	1953	Erie	Erie	PA	unknown
Sadie N.	1960	Girard	Erie	PA	clerk

Table 3.11: Demographic characteristics of 18 unrecorded speakers (they will only be displayed in the maps for minimal pairs and lexical / morphosyntactic variables)

ducted specifically for this dissertation (including speakers with no audio data) and from archival material.

The corpus contains 68 females and 38 males; thus the ratio of female to male speakers is 1.8:1. This ratio is nearly identical to the overall female to male ration of 1.7:1 in the ANAE corpus (Labov et al. 2006:28).

The map in Figure 3.2 shows the geographic location of all 88 speakers whose acoustic data will be analyzed in this dissertation, including both speakers from archival sources and speakers from interviews conducted for the dissertation.

3.6 Interview Procedure

3.6.1 Materials

For speakers who had enough time (ca. 30 minutes or more), the interview consisted of approximately 20 minutes of conversation and the formal methods. The conversation was targeted first to extract the necessary demographic information from the speaker, and then to ask them to describe their town (what the downtown is like, how it has changed) and talk about any other nearby towns or cities they frequently go to. Other topics that were often discussed at length include jobs, family, and hobbies. The conversations were thus similar in style to the interviews conducted for the ANAE, and are more properly characterized as dialectological interviews than sociolinguistic interviews.

The formal methods section of the interview consisted of a word list, a set of minimal pairs, and an acceptability judgment task. The word list consisted of 159 words, and was designed to provide a complete view of a speaker's vowel system. Extra tokens of words with vowels crucial to the present study (namely /o/, /oh/, /æ/, and /ow/) were included to ensure the reliability of mean values for these vowels (see Appendix C for the complete

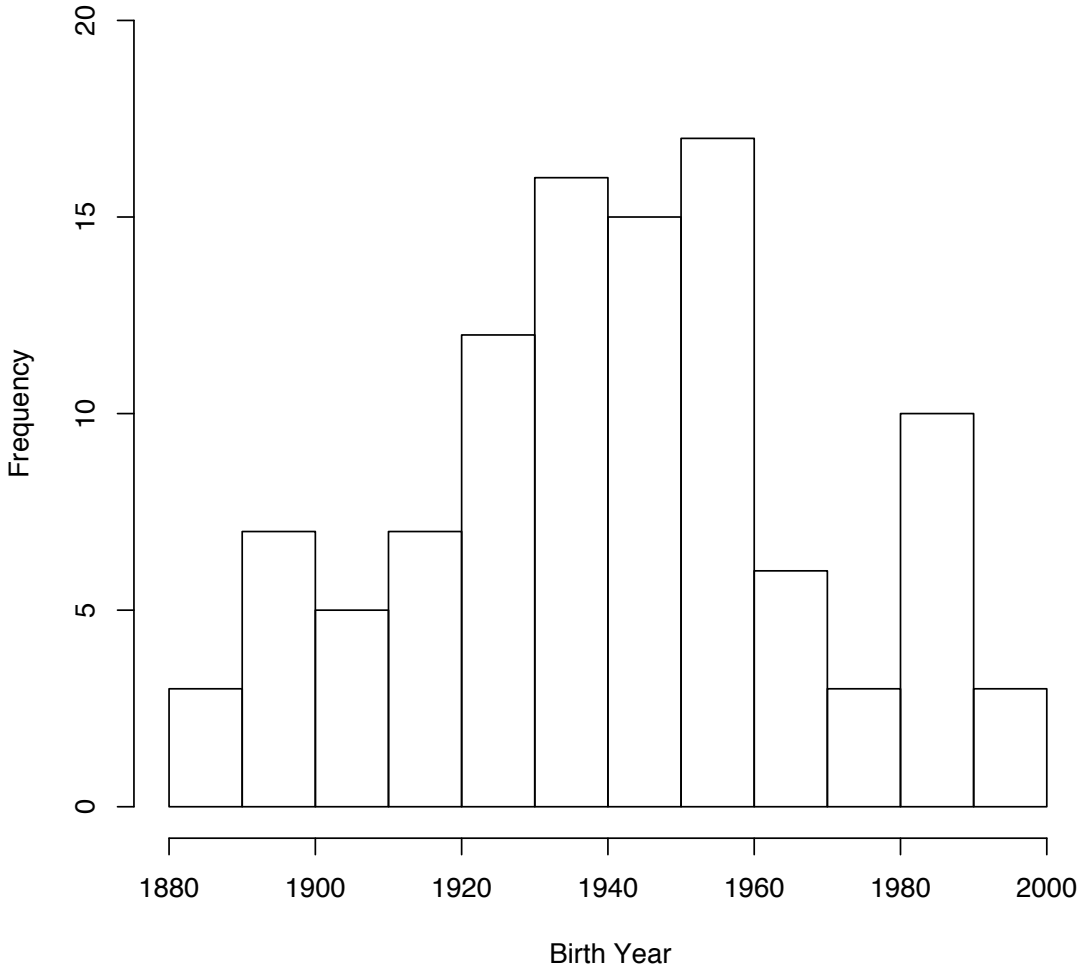

Figure 3.1: Histogram of birth year for all 106 speakers in the corpus (from both new interviews and archival sources)

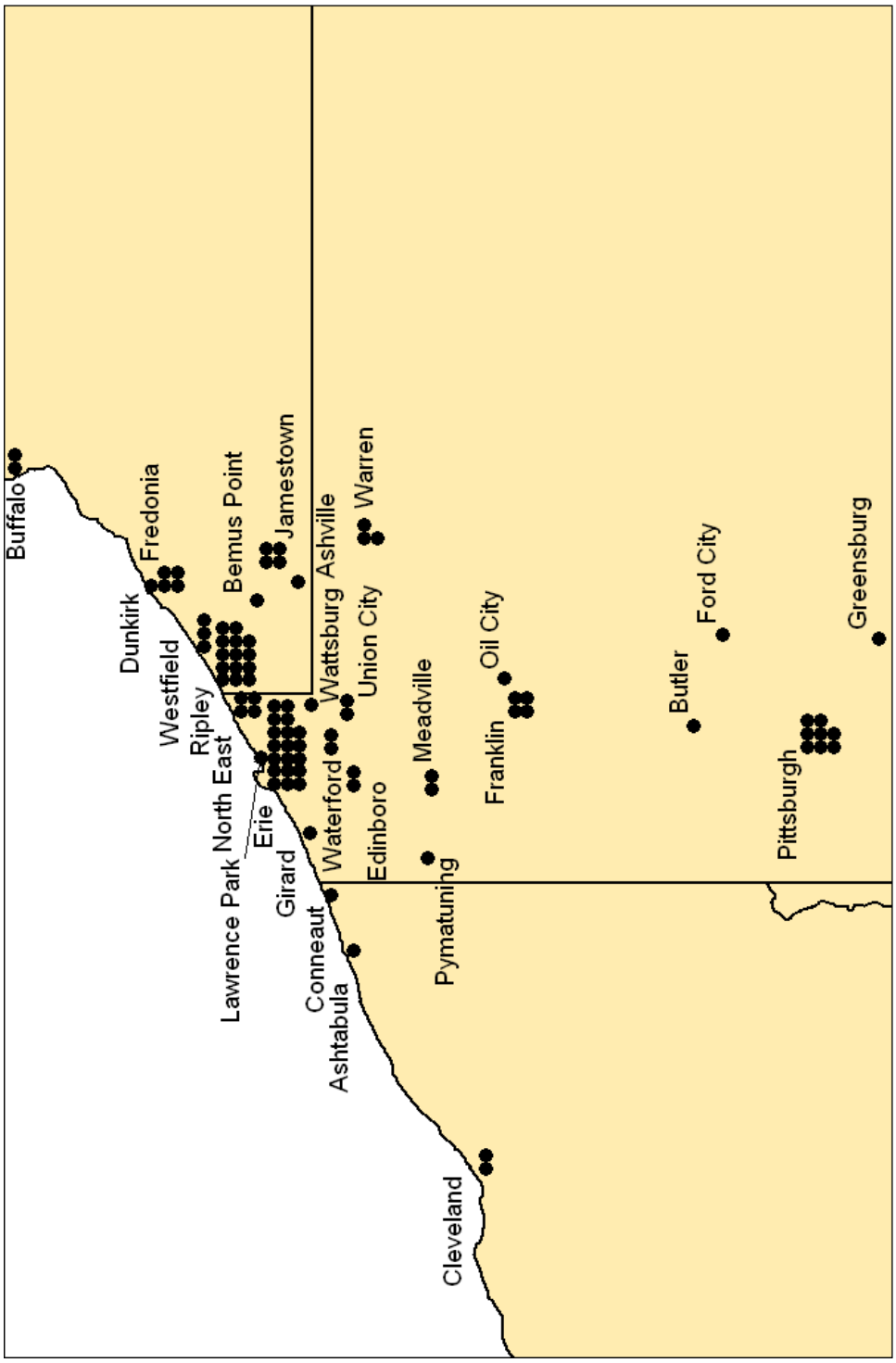

Figure 3.2: Locations of all 88 speakers whose acoustic data will be analyzed

word list). The minimal pair section contained 17 minimal pairs and 4 near-minimal pairs (see Appendix B for a complete list). Seven of the items involved the main distinction of interest, /o/ vs. /oh/, and two tested the contrast of /uw/ and /u/ before /l/. Two further near-minimal pairs examining the status of /ah/ were included, as well as one designed to determine the phonemic status of the vowel in *on*. The remaining pairs were included as filler material, and were not expected to vary within the geographic region under consideration. Finally, the acceptability judgment task consisted of 10 items designed to test the geographic extent of certain Midland lexical and morphosyntactic usages (see Appendix D for a complete list). For a few speakers who only had a limited amount of time, the interview consisted of at least the minimal pairs, and sometimes also the word list and sentence judgment task.

The audio data for the speakers from the DARE corpus consists of the "Arthur the Rat" reading passage and guided conversation. The methodology used for collecting this data is described in detail in the Introduction to Cassidy and Hall (1985). For this dissertation, interview data from three DARE speakers from Erie County will be analyzed, as well as "Arthur the Rat" data from 14 speakers (including the three from Erie County as well as 11 from the neighboring boundary regions.) The full text of the version of the "Arthur the Rat" reading passage used by DARE is included in Appendix E (Cassidy and Hall 1985:xliii).

The interviews conducted for the ANAE consist of a variety of elicitation techniques designed to encourage the speakers to say certain words, minimal pair tests, grammaticality acceptability ratings, and guided conversation (see Labov et al. (2006:29–35) for a complete description of the ANAE interview procedure). For this dissertation, the entire interviews with the speakers in Table 3.5 were transcribed and analyzed.

3.6.2 Equipment

The face-to-face interviews were conducted using an M-Audio Microtrack 24/96 solid state recorder with 16-bit quantization and a sampling rate of 44,100 Hz. A Sony ECM-717 lavalier microphone was attached to the speaker's shirt near the chin. The telephone interviews were conducted using a call collection interface provided by the Linguistic Data Consortium, and were separated into separate μ-law encoded audio files for each of the two channels. During the telephone interview, participants accessed a web site that contained the word list, minimal pairs, and sentence judgment task.

3.7 Transcription

The speech samples to be analyzed come from a wide variety of sources: interviews conducted specifically for this dissertation, interviews from archival sources (DARE, ANAE), word lists, and reading passages (DARE's "Arthur the Rat"). All of these types of speech materials were analyzed using the methodology of forced alignment and automatic vowel analysis described in Chapter 4. In this section, I will describe the procedure I followed to prepare the data for this type of analysis.

Before a sound recording can be processed by forced alignment, the speech must be transcribed orthographically. In order to complete the transcriptions quickly and efficiently, I developed a program that enabled me to transcribe the speech and non-speech sounds in each recording with a minimum of stopping and rewinding. The program, called `quickTrans`, is a collection of Python, Praat, and shell scripts, and can be downloaded freely as part of P2TK, the Penn Phonetics Toolkit.[4]

The general approach that `quickTrans` uses is to automatically segment the sound file into small chunks based on intensity levels of the signal. This is done by setting a mini-

[4]PT2K is available at *http://www.ling.upenn.edu/phonetics/p2tk/*.

mum intensity threshold level for a segment of speech to be considered part of an utterance, as well as a duration threshold for pauses. If a segment of speech has an intensity level continuously below the intensity threshold for a period of time longer than the minimum pause duration, then it is marked as a pause. The portions on either side of the pause with greater intensity are then marked as utterance chunks. These small chunks, which are intended to be short enough to fit into short-term memory, are then played sequentially by an audio player (Praat). Transcription can then proceed with a minimum amount of rewinding to repeat utterances.[5]

After some experimentation, the intensity threshold for pauses was set to 1/3 of a standard deviation less than the mean, and the minimum pause duration was set to 200 msec. In other words, if there is a segment of the audio file which has an intensity level consistently less than 1/3 of a standard deviation below the mean intensity level of the sound file for greater than 200 msec, then it is marked as a pause between utterances. These configuration values worked well for most of the recordings in my corpus. However, they needed to be modified for some recordings with a large amount of overlapping speech, constant background noise, or heavily unbalanced intensity levels between two interview partners (e.g., because the microphone was much closer to one interlocutor than the other). In such cases, the `quickTrans` configuration variables were modified until they produced utterance chunks that were mostly short enough in duration to fit into short-term memory and that did not omit segments of speech data.

This approach used by `quickTrans` is similar to the `AutoSegmenter` tool from the LDC (Glenn and Strassel 2008). It enables quick transcription, where the goal is simply to "get the words right" as quickly as possible. The LDC reports transcription rates of approximately seven to ten times real-time using `AutoSegmenter` while following

[5]If the transcribed is not able to transcribe an entire utterance on the first pass, the `quickTrans` interface does enable the entire utterance to be repeated. However, more complex manipulations, such a replaying part of an utterance or listening to previous utterances must be done in an external sound editor.

the quick transcription guidelines described in LDC (2004). For my own transcriptions of my dissertation corpus, I was able to transcribe most files at a rate of around five times real-time. This fast rate of transcription was partially due to the fact that I omitted all information from the transcription that was not necessary for the purposes of forced alignment and automatic vowel analysis. Specifically, the transcriptions include no punctuation and are all in lower case. On the other hand, certain transcription conventions were followed in order to make the forced alignment as accurate as possible (see Evanini et al. (2009b) for a more complete description of these transcription practices and how they affect the output of forced alignment). These included:

- **Disfluencies**: Partial word disfluencies, such as the false start of *sch-* for *school* are transcribed with a hyphen representing the part of the word that was not spoken. Pause fillers, such as *um* and *uh*, are always transcribed.

- **Non-speech sounds**: The forced alignment system recognizes five symbols for non-speech sounds: {BR} for breath, {CG} for cough, {LG} for laughter, {LS} for lip smack, and {NS} for background noise. These were transcribed as such when they were loud enough to be included in the utterance chunks produced by `quickTrans`.

- **Unknown words**: Words that are uttered in the audio file but that do not have a corresponding transcription in the forced alignment system's pronouncing dictionary are problematic. If no pronounciation is available, then the system can not include them in the alignment; this also leads to sub-optimal alignments for the words in the transcription surrounding the missing word. Therefore, after a post-transcription check with the pronouncing dictionary, all words in the transcription that were missing from the dictionary were manually provided with pronunciations.

- **Multiple speakers**: For audio files with multiple speakers, it is necessary to annotate which speaker uttered each word. Without this, automatic vowel analysis would

be impossible, since the identity of the speaker for each token would be unknown. The convention I used in my transcriptions was to prepend a single capital letter corresponding to the speaker that produced each utterance for every audio file with speech data from more than one speaker.[6]

For example, a portion of the transcription for the interview with Dottie A. is reproduced below. It illustrates the transcription of a partial word disfluency (*sch-* for *school*), filled pauses and backchannels (*uh* and *uh huh*), non-speech sounds ({LG} for laughter, {BR} for breath, and {NS} for background noise) and input from multiple speakers (Speaker `A` is Dottie A. and Speaker `B` is the interviewer (me)).

```
A i taught in panama new york
B uh huh A and uh
A then i came back to erie
A and when i got a teaching job again
A i taught just middle sch- middle school and junior high
B and how did you enjoy that job
A well i liked it i liked the subject of home economics
A but i
A wasn't too crazy about the discipline i had to {LG}
A {BR}
A to uh manifest in order to
A keep the children
{NS}
A so that they would
A be getting some kind of an education
```

3.8 Vowel Measurements

Vowel formants were extracted automatically for all speakers in the corpus (except for the two speakers from the SWV corpus, as described above) according to the procedures

[6]Because all interviews were transcribed in lower case, this convention never produced any ambiguities with words from the transcription.

described in Chapter 4. A collection of Python scripts were written to implement the automatic vowel analysis techniques developed for this dissertation. They are included in the program `extractFormants`, also available as part of the P2TK package.

The `extractFormants` configuration variables were set so that only vowels longer than 50 msec in duration (as determined by the output of the forced alignment procedure) were measured. In total, 113,245 F1 and F2 measurements were extracted (vowel tokens from the interviewer and other speakers in the recording that will not be analyzed are excluded from this figure). This number includes vowels with all three levels of lexical stress: primary, secondary, and none, as well as vowels in all segmental environments.

After the vowel formants were extracted, they were normalized on a per-speaker basis to reduce the effects of individual variation in vocal tract length, according to the log-mean algorithm in Nearey (1977) (see also Labov et al. (2006:39–40) for a concise summary of the procedure). The group log mean value from the ANAE, 6.896874, was used for this corpus as well, since it was calculated from a larger number of speakers.

In an attempt to remove outliers due to errors introduced by the automatic formant measurement process, the automatic measurements were compared to the database of manual measurements taken for the ANAE. Any normalized automatic measurement that was outside the range of all of the normalized hand measurements for each vowel class was excluded from the set of automatic measurements. This approach is justified, since it is quite unlikely that such measurements represent accurate formant values for the vowel. Since the ANAE database is both larger than the database for this dissertation and drawn from a wider variety of dialects, it would be expected to exhibit a wider range of formant values for each vowel class. Thus, any measurement outside of these ranges can be treated as erroneous. A total of 1,282 automatic measurements were excluded based on this metric, approximately 1% of the total database. Afterwards, 111,963 vowel measurements remained for analysis.

As an example of the usefulness of excluding automatic measurements outside the

bounds of the manual ANAE measurements, consider the vowel plot of /o/ and /oh/ for Dan R. from Erie shown in Figure 3.3. A single measurement from the /o/ class is a clear statistical outlier. This token of *got*, shown in the lower-left corner of the vowel plot, was provided with an F1 measurement of 1247 Hz and an F2 measurement of 2115 Hz by the automatic vowel analysis procedure described in Chapter 4.[7] However, this measurement is clearly an error, since the articulation of a vowel with such a high F1 value would be physically impossible for this speaker. Additionally, manual inspection of this token shows that the correct (unnormalized) F1 and F2 measurements should be around 501 Hz and 1206 Hz, respectively. Clearly, the automatic formant prediction procedure erred in this case by substituting F2 for F1 and F3 for F2.

The exclusion procedure described above applies to this token of /o/ in *got* for both the F1 and the F2 values. The normalized F1 value for this measurement is 1520 Hz, and the maximum normalized F1 value for all manual measurements of /o/ in the ANAE database is 1264. Additionally, the normalized F2 value for this automatic measurement is 2578, and the maximimum normalized F2 value for /o/ in the ANAE is 1915. Thus, both the F1 and F2 measurements for this token of *got* are outside the range of manual measurements for /o/, and this token is correctly labeled as an error by this procedure and excluded from consideration.

The exclusion of such outliers caused by measurement errors is useful, because such gross errors can have a disproportionately large effect on the mean values. For example, Dan R.'s F1 and F2 mean values for 56 tokens of /o/ before the exclusion of outliers are 717 Hz and 1361 Hz, respectively. After this single token of *got* with the gross measurement error is excluded, the mean values change to 704 Hz and 1338 Hz. This change in the mean values is not so drastic for Dan R., because his number of /o/ tokens is quite high. However,

[7]These values represent the original unnormalized measurements produced by the system, whereas the plot in Figure 3.3 shows the measurements after normalization. That is why the stated values do not correspond exactly to the coordinates in the figure.

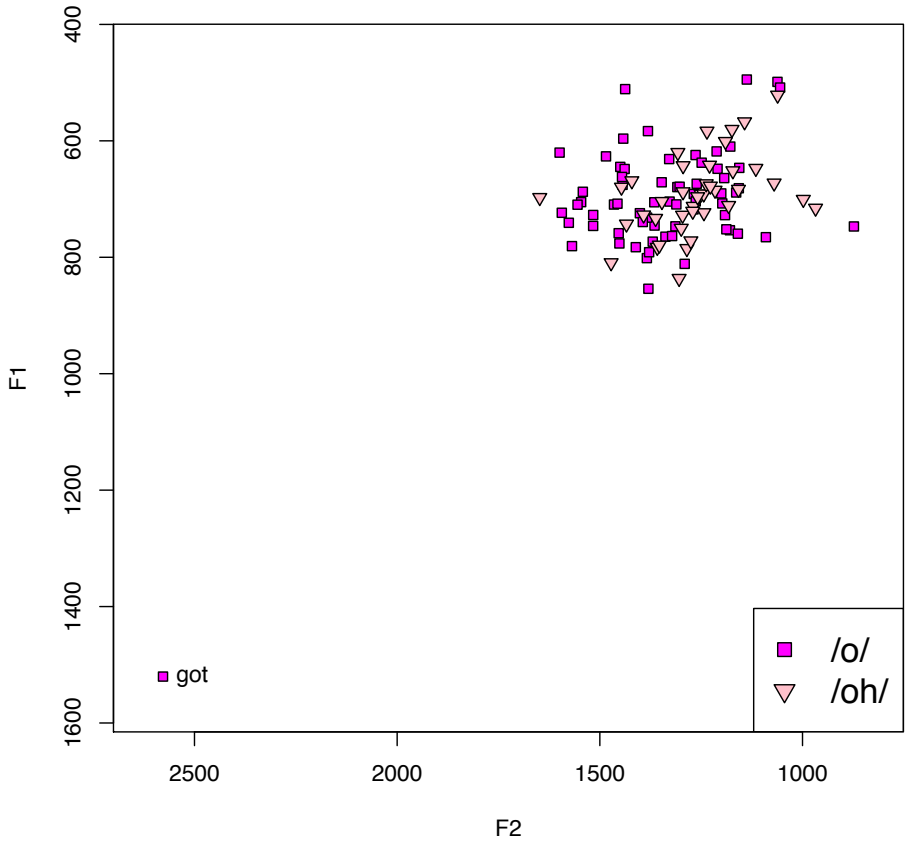

Figure 3.3: Automatic formant measurements for the vowels /o/ and /oh/ for Dan R. from Erie. The token of *got* in the lower-left corner is a gross measurement error that should be excluded from a vowel analysis.

a single gross measurement error such as this one can cause the mean values to shift by a few hundred Hz when the number of tokens is small. This point becomes important for the DARE speakers whose /o/ and /oh/ values are analyzed from the "Arthur the Rat" reading passage. As Section 6.7.3 will describe, each speaker produces about 14 tokens of /o/. For these speakers, the exclusion of gross errors is necessary in order to obtain reliable mean values.

After gross errors were excluded by this procedure, a subset of the remaining 112,087 vowel measurements were then selected when calculating vowel means for individual speakers in Chapters 5 through 7. The following list provides details about which tokens were excluded:

- vowels before /l/ and /r/
- vowels after /w/ and /y/
- vowels after obstruent+liquid onset clusters
- tokens of /i/, /e/, /æ/, and /aw/ before nasals
- vowels with secondary stress and unstressed vowels (as indicated by the phonemic transcription in the CMU pronouncing dictionary)
- vowels in a set of high-frequency function words that often undergo reduction: *and, but, for, he, he's, huh, I, I'll, I'm, is, it, it's, its, my, of, oh, she, she's, that, the, them, then there, they, this, uh, um, up, was, we, were, what, you*

The first four sets of exclusions based on the neighboring segmental environments correspond to the exclusions that the ANAE authors made before means were calculated (Labov et al. 2006:77). The exclusion of vowels not marked with primary lexical stress and vowels from words in the list of stop words represents an attempt to reproduce the ANAE's

selection criteria. Under their approach to vowel analysis, the annotators only measured vowels bearing primary lexical stress. Additionally, most words chosen for analysis also bore primary phrasal stress (Labov et al. 2006:37). Focusing only on these tokens for vowel mean measurements reduces the centralizing effect of vowel reduction and presents a more accurate view of the phonetic targets for each vowel.

After these exclusions were applied, the total number of remaining vowel measurements was 44,599. Unless stated otherwise, all maps and figures displayed below were generated from this smaller subset of 44,599 vowels.