Chapter 3

The Philadelphia Neighborhood Corpus Data

In this chapter, I'll briefly describe the data used in this dissertation from the Philadelphia Neighborhood Corpus. I'll try not to be overly redundant with descriptions which are already in press (Labov et al., 2013; Evanini, 2009), but I have enriched the data to some extent, which requires some explanation.

3.1 The Philadelphia Neighborhood Corpus

The Philadelphia Neighborhood Corpus [PNC] contains sociolinguistic interviews carried out in Philadelphia between 1972 and 2012 (at the time of this writing). These interviews were carried out as part of coursework for Ling560 'The Study of the Speech Communuty.' Each year the course was taught (annually from 1972 to 1994, every other year from then on), students formed into research groups, and selected a city block on which to base their study. For more information on Ling560 and the neighborhoods which have been studied, see Labov et al. (2013). The total Ling560 archive contains interviews with 1,107 Philadelphians. Not taking into account the interviews collected in the 2012-2013 academic year, interviews with 379 speakers have been transcribed by undergraduate research assistants, and included in the PNC.

3.1.1 Forced Alignment and Vowel Extraction (FAVE)

The audio recordings and transcriptions of the interviews were then processed by the Forced Alignment and Vowel Extraction (FAVE) suite (Rosenfelder et al., 2011). As the name would suggest, there are two steps to the FAVE analysis. First is forced alignment, which aligns words and phones to the audio. The acoustic models for FAVE come from the Penn Phonetics Lab Forced Aligner [p2fa] (Yuan and Liberman, 2008), with some extra procedures added to account for overlapping speech. With the forced alignment, we can identify where in the audio a particular vowel begins and ends.

The second step is automated vowel formant analysis, an approach first attempted by Evanini (2009). The errors involved in LPC formant analysis are frequently catastrophic, and it was for this reason that the authors of the Atlas of North American English concluded that automated formant analysis was not feasible at the time (Labov et al., 2006). For example, for a vowel like /iy/, in which there is a large distance between F1 and F2, an LPC analysis using 12 poles might erroneously detect a formant between F1 and F2, providing formant estimates with an F2 which is too low. On the other hand, for a vowel like /ɔ/, where F1 and F2 are very close, an LPC analysis using 6 poles might not differentiate F1 and F2, and would return what is actually F3 as F2. In practice, errors like these have been handled by a researcher visually comparing LPC estimates to the spectrogram, and to their personal prior expectations for what the formants of this particular vowel ought to be, adjusting the LPC parameter settings accordingly.

What the FAVE suite does is replace a researcher's prior expectations with quantitative priors from the Atlas of North American English. The vowel class we are trying to measure is given by the forced alignment, meaning that the acoustic data is labeled. Drawing from the ANAE, we can establish certain expectations for the formant measurements we can expect for the specific label. For 4 different LPC parameter settings (6, 8, 10, and 12 poles) we extract the estimates for F1 and F2 frequencies and bandwidths, and compare these to our priors from the Atlas of North American English using the Mahalanobis distance. The LPC parameter setting with the smallest Mahalanobis distance is taken to be the winner. This process is repeated for every vowel in the speaker's interview.

We found that even after we chose LPC parameter settings based on comparison to the Atlas of North American English, there were still a small number of gross errors in the data. We guessed that this may be due to the fact that priors based on the entire ANAE may not be the most appropriate priors for each individual speaker. The most appropriate prior expectation for how a speaker ought to pronounce a vowel is actually how *that speaker* usually pronounces that vowel. Having eliminated most gross errors from the speaker's vowel measurements through comparison to the ANAE data, we could now generate reasonable speaker specific expectations for each vowel.[1] As a second step, then, FAVE iterates through all of the vowel measurements again, this time comparing the different LPC settings to the speaker specific vowel distributions. This second step, in addition to vowel specific heuristics for selecting a measurement time point,[2] eliminates almost all gross errors.

For the time being, the FAVE system can only be assured to give high quality results for North American English (because of the reliance on ANAE priors), and only for an aligned corpus using the CMU dictionary transcriptions (which is what p2fa and FAVE-align use). However, extending the method to any given dialect or language is conceptually trivial. First, a certain number of high quality hand measurements for each vowel in the dialect needs to be collected. The ANAE is a large database, but the sample size necessary to establish the first pass priors need not be as large. Even relatively small numbers of measurements, say 10 to 20 per vowel, ought to be sufficient, since the goal of the priors is not to provide an overly precise estimate for each vowel, but rather just to weed out the grossest errors. With these priors collected, the implementation of FAVE would need to be changed to not make specific reference to the CMU labels. Of course, FAVE-extract is dependent on having an alignment, which for now is based on the models from p2fa. However, for other dialects or languages, there are other trainable aligners available, like the Prosodylab-Aligner (Gorman et al., 2011).

[1] This was my own substantive contribution to the FAVE suite.

[2] These were explored and implemented by Ingrid Rosenfelder.

3.2 Enrichment of Contextual Information

FAVE-extract provides two different kinds of output file types: the Plotnik file formant (Labov, 2006), and a tab delimited file. For the purpose of my dissertation, the data available in these outputs was not entire sufficient. With regards to the contextual coding, they indicate the place, manner and voicing of the following segment, the nature of the syllable coda, and how many following syllables in the word, and the quality of the preceding segment. The actual label for the preceding or following segment is not included, nor the transcription of the word, or any contextual information across word boundaries. Some of this information is crucial for my analyses, so I enriched the information available from the original FAVE-extract output. Using the time stamp of the measurement point, I scanned the Praat TextGrids which served as the input for FAVE-extract and tried to identify the vowel that corresponded to a particular measurement. This step was complicated by the fact that some vowels were not measured within the boundaries provided by the FAVE-align, but rather within the boundaries of the preceding segment. Any vowel which could not be programmatically located in a TextGrid were discarded. In addition, one speaker's TextGrid could not be located in the corpus, and their data has also been excluded from this dissertation.

Once locating the correct vowel in the TextGrid, I extracted the following information for the vowel.

(3.1) the full CMU transcription for the word the vowel is located in.

(3.2) the preceding segment, disregarding word boundaries.

(3.3) the following two segments, disregarding word boundaries.

(3.4) the location of the vowel in the word, coded as

  (a) word initial

  (b) word final

  (c) coextensive with word boundaries (e.g. *I*)

  (d) word internal

These additional pieces of information were crucial for most analyses in this dissertation. For example, for word final vowels, knowing the following segment was crucial for investigating whether the conditions on certain phonetic changes applied at the phrase or word level. Also, with the full CMU transcription of the word, I was able to apply simple syllabification algorithms which allowed me to, for example, compare open and closed syllables on the conditioning of /ey/, and to identify which following /t/ and /d/ were flapped when looking at /ay/.

3.3 Total Data Count

The Ling560 fieldworkers have visited a relatively racially diverse set of neighborhoods. However, the vast majority of speakers included in the PNC so far are of White European descent. It would be a mistake to treat data drawn from African American Philadelphians and White Philadelphians as being drawn from one unified speech community. The two social groups clearly form separate, but mutually influencing, speech communities (Labov et al., 1986). In fact, Henderson (2001) found that listeners could correctly identify White and African American Philadelphians' race simply from a recording of them counting from 1 to 20. Despite the facts that the mutual influence of these two dialects on each other is so interesting, and that the White Philadelphian dialect is spoken now by a numerical minority of all Philadelphians, the nature of the data available at the moment constrains me to look exclusively at White speakers.

Taking into account that I will only be examining the data from White Philadelphians, that one speaker had to be excluded because I could not locate their TextGrid, and that some vowel measurements had to be excluded because the vowel could not be programmatically located in their TextGrid, I will be working with 735,408 vowel measurements from 308 speakers. Figure 3.1 plots a histogram of how many vowel measurements are available from each speaker.

3.4 Normalization

All of the data were normalized to formant intrinsic z-scores (i.e. Lobanov Normalization) (Adank et al., 2004). In this dissertation, I will be using the z-score measure directly, rather than rescaling

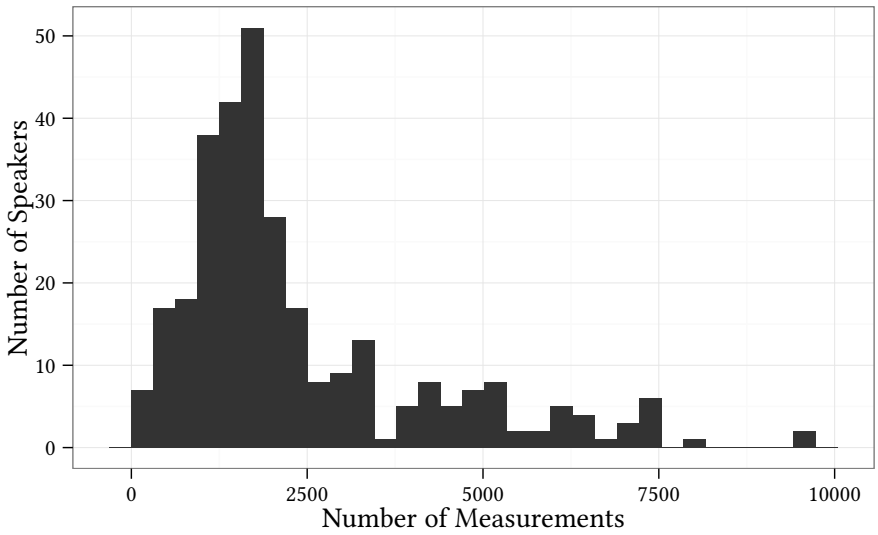

Figure 3.1: Histogram of how many vowel measurements are drawn from each speaker.

it to a hertz-like measure.

3.5 Choice of Time Dimension

There are a number of different possible time dimensions available from the corpus (see Sankoff (2006) for an overview of real and apparent time). I could, for example, use a strictly real time measure, and evaluate the phonetic changes I investigate against their year of interview. There are also two different apparent time measures available, speakers' Age and Date of Birth. Figure 3.2 plots the year of interview and the age of the speaker, while Figure 3.3 plots the date of birth of the speaker, and their age at the time of the interview. What should be clear from Figure 3.3 is that any result obtained using a speaker's date of birth is going to be very similar if the speaker's age was used instead. The high correlation between speaker's age and date of birth is simply due to the facts of human lifespan, and that the fieldwork has covered 40 years.

However, it is possible to compare statistical models that use each kind of time dimension to see which has the best predictive power. Labov et al. (2013) did this by comparing the r^2 of

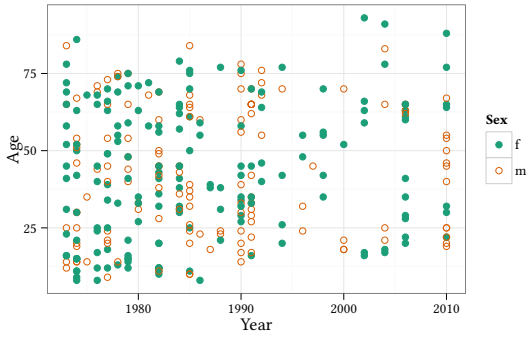

Figure 3.2: Year of interview, and age of speaker.

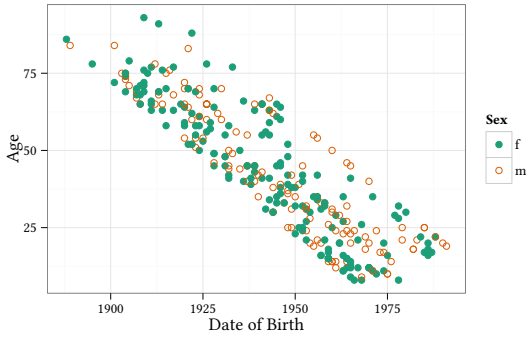

Figure 3.3: Speakers' date of birth, and age at time of interview

each, and I will briefly replicate that analysis here. Figures 3.4 to 3.6 plot the relationship between the normalized F1 of pre-voiceless /ay/ and the three possible diachronic dimensions (year of interview, at at interview, and date of birth).

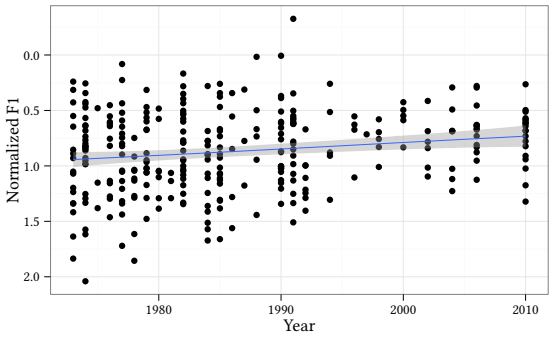

Figure 3.4: Relationship between /ay/ raising and year of recording.

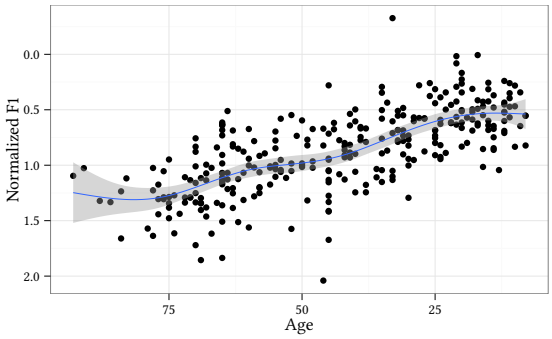

Figure 3.5: Relationship between /ay/ raising and speaker's age.

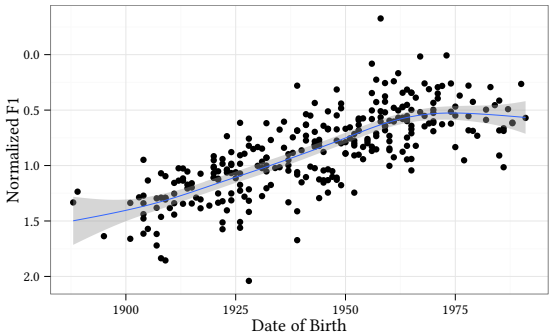

Figure 3.6: Relationship between /ay/ raising and speaker's date of birth.

I fit three generalized additive models to predict the F1 of pre-voiceless /ay/ using cubic regression splines for each sex.[3] Table 3.1 displays the r^2 and the Akaike Information Criterion [AIC] for each model. The model predicting pre-voiceless /ay/ F1 using speakers' date of birth has the highest r^2 and the lowest AIC, suggesting that this ought to be the preferred model.

Predictor	r^2	AIC
Year	0.03	247
Age	0.49	53
Date of Birth	0.59	-11

Table 3.1: Model comparisons for using different time dimensions to predict pre-voiceless /ay/ height.

Pre-voiceless /ay/ raising exhibits one of the two patterns of change in Philadelphia that Labov et al. (2013) identified (linear incrementation). Just to make sure that date of birth is also the best diachronic dimension for the other pattern of change (reversal), I fit three models predicting the fronting and raising of /aw/ using year of the recording, speakers' age, and speakers' date of birth. The r^2 and AIC for these models are displayed in Table 3.2. The model using date of birth again has the highest r^2 and lowest AIC, suggesting that for the changes which reversed course, date of birth is also the best diachronic dimension to use. The fact that the r^2 for the best /aw/ is much smaller than the r^2 of the best /ay/ model is probably due to the fact that /aw/ is more highly differentiated along social dimensions, as Labov et al. (2013) found when they took into account speakers' level of education.

Predictor	r^2	AIC
Year	0	454
Age	0.11	423
Date of Birth	0.13	417

Table 3.2: Model comparisons for using different time dimensions to predict /aw/ raising and fronting.

Given that date of birth has the best predictive power for both /ay/ and /aw/, which themselves exemplify the two major patterns of change I investigate in this dissertation, I'll be using date of

[3] The formula was `gam(F1.n ~ s(X, bs = ''cs'', by = Sex))`.

birth as the diachronic dimension throughout the dissertation.

Bibliography

Adank, Patti, Roel Smits, and Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116:3099. URL `http://link.aip.org/link/JASMAN/v116/i5/p3099/s1&Agg=doi`.

Evanini, Keelan. 2009. The permeability of dialect boundaries: A case study of the region surrounding Erie, Pennsylvania. Ph.d., University of Pennsylvania.

Gorman, Kyle, Jonathan Howell, and Michael Wagner. 2011. Prosodylab-Aligner: A tool for forced alignment of laboratory spech. In *Proceedings of Acoustics Week in Canada, Quebec City*, 4–5.

Henderson, Anita Louise. 2001. Is your money where your mouth is? Hiring managers' attitudes toward African-American Vernacular English. Ph.d., University of Pennsylvania.

Labov, William. 2006. Plotnik 08. Retrieved September 2006 http://www.ling.upenn.edu/∼wlabov/Plotnik.html.

Labov, William, Sherry Ash, and Charles Boberg. 2006. *The Atlas of North American English.* New York: Mouton de Gruyter.

Labov, William, David Graff, and Wendell A. Harris. 1986. Testing listeners' reactions to phonological markers. In *Diversity and Diachrony*, ed. David Sankoff, 45–58. Philadelphia: John Benjamins.

Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One hundred years of sound change in Philadelphia: Linear Incrementaion, Reversal, and Reanalysis. *Language* 89:30–65.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite. URL `http://fave.ling.upenn.edu/`.

Sankoff, Gillian. 2006. Gillian Sankoff. Age: Apparent time and real time. In *Encyclopedia of language and linguistics*. Oxford: Elsevier.

Yuan, Jiahong, and Mark Liberman. 2008. Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics*.