Speech Rate, Pause, and Linguistic Variation: 
An Examination Through the Sociolinguistic Archive and Analysis Project

by

Tyler S Kendall

Department of English
Duke University

Date: ______________________

Approved:

___________________________
Walt Wolfram, Supervisor

___________________________
Agnes Bolonyai

___________________________
Ronald Butters

___________________________
Erik Thomas

Dissertation submitted in partial fulfillment of 
the requirements for the degree of Doctor 
of Philosophy in the Department of 
English in the Graduate School 
of Duke University

2009
ABSTRACT

Speech Rate, Pause, and Linguistic Variation: 
An Examination Through the Sociolinguistic Archive and Analysis Project

by

Tyler S Kendall

Department of English
Duke University

Date: _________________________

Approved:

___________________________
Walt Wolfram, Supervisor

___________________________
Agnes Bolonyai

___________________________
Ronald Butters

___________________________
Erik Thomas

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of English in the Graduate School of Duke University

2009
Abstract

Recordings of speech play a central role in the diverse subdisciplines of linguistics. The reliance on speech recordings is especially profound in sociolinguistics, where scholars have developed a range of techniques for eliciting and analyzing natural talk. Despite the focus on naturalistic speech data, sociolinguists have rarely focused explicitly on the management (e.g. organization, storage, accessibility, and preservation) of their data, and this lack of focus has had consequences for the advancement of the field. At the same time, the interviews that sociolinguists labor so hard to obtain are often barely mined for their full potential to further our understanding of language. That is, sociolinguists often focus on a handful of phonological and/or morphosyntactic variables to the exclusion of so many other features of speech. The present work both addresses the management of sociolinguistic data and, through an innovative approach to speech data management and analysis, extends the sociolinguistic lens to include the lesser-examined realm of variation in sequential temporal patterns of talk.

The first part of this dissertation describes the Sociolinguistic Archive and Analysis Project (SLAAP), a web-based digitization and preservation initiative at North Carolina State University. SLAAP, which I principally have designed and developed, is more than an archive; it has actively sought to explicate approaches to spoken language data management and to enrich spoken language data through the development of analytic tools designed specifically for sociolinguistic analysis. This dissertation begins by situating SLAAP within the history of data management practices in the field of
sociolinguistics. It then provides an overview of many of SLAAP’s features, discussing in particular the transcript model that enables most of its analytic and presentational capabilities.

The second part of this dissertation takes advantage of SLAAP’s data model and the extensive language data accumulated within its archive to examine variation in speech rate and silent pause duration by North American English speakers of four ethnicities in North Carolina, Ohio, Texas, Washington, DC, and Newfoundland. This work brings a wide range of previous research from different areas of sociolinguistics, psycholinguistics, and corpus linguistics to bear on an array of quantitative analyses, demonstrating that speech rate and pause exhibit meaningful variation at the social level at the same time as they are also constrained by cognitive and articulatory processes.

Specifically, pause and speech rate are shown to vary by region, ethnicity, and gender – albeit not in mono-directional ways – although other factors arise as significant, including, for speech rate, a strong effect of utterance length as well as a number of interactional or discourse-related factors, such as the gender of the interviewer and the number of participants in the speech event. A number of the examinations undertaken relate sociolinguistic conceptions of style to language production and cognitive processes, including a quantitative analysis of sequential temporal patterns as paralinguistic cues to attention to speech, performativity, and the realization of phonological and morphosyntactic variables. Through this analysis sociolinguistic data and findings are brought to bear on a tradition of psycholinguistic investigations with the hope to benefit both, often disparate, areas of research.
To

Herbert Nathan Putnam III

for completing my two families.
Contents

Abstract ........................................................................................................................................... iv

List of Tables ................................................................................................................................... xi

List of Figures ................................................................................................................................... xiii

Acknowledgements ........................................................................................................................... xvi

Part 1. Reconsidering speech as data: The Sociolinguistic Archive and Analysis Project . 1

1. Introduction ................................................................................................................................. 1

1.1. Rationale and overview of the present work ................................................................. 1

1.2. A short history of data within sociolinguistics .......................................................... 6

1.3. How does speech constitute data? ................................................................................. 10

1.4. Data vs. metadata ............................................................................................................ 16

1.5. Treating speech as data in the 21st century ............................................................... 20

1.6. In closing: Managing speech data for sociolinguistic analysis ................................ 24

2. Extending the utility of the archive ......................................................................................... 27

2.1. The Sociolinguistic Archive and Analysis Project .................................................... 27

2.2. Features of the SLAAP software .................................................................................. 28

2.3. Some benefits of explicit sociolinguistic data management work ......................... 34

2.4. SLAAP’s data model ....................................................................................................... 41

2.5. In closing: Strengthening our relationship with our data ........................................ 44

3. Implementing a databased transcription model ................................................................. 46

3.1. Transcription as theory and practice ........................................................................... 46

3.2. What constitutes a transcript line? .............................................................................. 50

3.3. Transcribing for SLAAP .............................................................................................. 60
3.4. Visualizing speech data.................................................................66
3.5. In closing: Transcription beyond text, annotation beyond transcription ........74

Part 2. Extending the variationist lens to speech rate and pause............................76

4. Introduction to Part 2 ...............................................................................76
4.1. Directions for extension: Investigating pause and speech rate...................76
4.2. The data.............................................................................................78
4.3. The tools............................................................................................84
4.4. Caveats for the present exploration........................................................87
4.5. In closing: A corpus-like approach.........................................................90

5. Toward a quantitative sociolinguistic analysis of pause..................................91
5.1. Introduction .........................................................................................91
5.2. Research on pause...............................................................................94
5.3. Defining and measuring silent pause.....................................................101
5.4. Analysis and discussion .....................................................................109
5.5. Pause type and filled pause adjacency on silent pause durations ............118
5.6. In closing: Variation in pause...............................................................126

6. Toward a quantitative sociolinguistic analysis of speech rate .......................129
6.1. Introduction .........................................................................................129
6.2. Defining speech rate terminology.........................................................129
6.3. Research on speech rate ......................................................................130
6.4. Automating speech rate measurements................................................136
6.5. Analysis and discussion .....................................................................139
6.6. In closing: Variation in speech rate .......................................................153

7. The relationship between speech rate and pause ........................................155
7.1. Are speech rate and pause related? ................................................................. 155
7.2. Modeling speech rate and pause duration .......................................................... 156
7.3. In closing: Variation in speech rate and pause duration ................................. 165
8. Style and register variation in speech rate and pause duration ................. 168
   8.1. Introduction ........................................................................................................ 168
   8.2. Pause and speech rate and stylistic variation .................................................... 169
   8.3. In closing: Register and style variation in pause and speech rate .......... 176
9. Pause and speech rate in interaction ................................................................. 178
   9.1. Introduction ........................................................................................................ 178
   9.2. Some problems with Chapter 5, 6, and 7 ......................................................... 178
   9.3. Interlocutor effects on pause and speech rate ............................................... 181
   9.4. A case study: Who is interviewing EH? .......................................................... 191
   9.5. A case study: Carissa is interviewing whom? ................................................. 194
   9.6. In closing: Interactional variables vs. sociolinguistic variables? ............. 203
10. Speech rate and pause and “traditional” sociolinguistic variables ............ 206
    10.1. Introduction ...................................................................................................... 206
    10.2. Performance and production ......................................................................... 207
    10.3. The “Henderson graph” redux: Paralinguistic cues to variable realization ... 210
    10.4. A case study: Language and identity in Petty Harbour, Newfoundland .......... 214
    10.5. In closing: sequential temporal patterns and variable realization .......... 232
Part 3. Towards the future ......................................................................................... 233
11. Conclusion ............................................................................................................. 233
    11.1. In final closing ................................................................................................ 233
    11.2. Summary of findings (from Part 2) ................................................................. 234
11.3. For the future ........................................................................................................238

Appendix: Data summary for primary 104 speakers used in Part 2.......................240

References..................................................................................................................244

Biography....................................................................................................................260
List of Tables

Table 3.3.1. SLAAP transcription conventions, special characters ........................................ 64
Table 4.2.1. Ethnicity, gender, and region of speakers used in analyses .......................... 81
Table 5.4.1. Pause data by ethnicity .................................................................................. 111
Table 5.4.2. Pause data by gender .................................................................................... 112
Table 5.4.3. Pause data by region .................................................................................... 112
Table 5.4.4. Mixed-effect model for all pause duration data ......................................... 115
Table 5.5.1. Pause duration summaries by pause type ..................................................... 119
Table 5.5.2. Pause duration summaries by adjacency to filled pauses ......................... 121
Table 5.5.3. Mixed model for pause, with pause type and filled pause factor groups .... 124
Table 5.5.4. Mixed model for all pause duration data ..................................................... 115
Table 5.5.5. Mixed model for pause duration summaries by pause type ....................... 121
Table 5.5.6. Mixed model for pause, with pause type and filled pause factor groups .... 124
Table 6.5.1. Speech rate data by ethnicity ....................................................................... 142
Table 6.5.2. Speech rate data by gender .......................................................................... 142
Table 6.5.3. Speech rate data by region ........................................................................... 143
Table 6.5.4. Mixed-effect model for all speech rate data .............................................. 144
Table 6.5.5. Mixed-effect model for speech rate, examining the gender + ethnicity interaction .......................................................... 147
Table 6.5.6. Mixed-effect model for speech rate including utterance length .............. 152
Table 7.2.1. Fixed-effect model for influences on speech rate ...................................... 159
Table 7.2.2. Fixed-effect model for influences on pause log-duration ......................... 162
Table 8.2.1. Comparison of pause and speech rate with diagnostic variables in two speech settings for the County Commissioner from Roanoke Island, NC .......... 170
Table 8.2.2. Comparison of pause and speech rate with diagnostic variables in two speech settings for the Town Manager of Princeville, NC ........................................ 171
Table 8.2.3. Comparison of pause and speech rate with diagnostic variables in two speech settings for the Mayor of Princeville, NC ...................................................... 172
Table 9.3.1. Minor and non-significant differences between subset and main data ....... 182
Table 9.3.2. Fixed-effect model for influences on speech rate, including interviewer
gender.................................................................................................................. 184
Table 9.4.1. Interviewers and data for EH.................................................................. 192
Table 9.5.1. Median pause duration and median speech rate for the twelve African
American teenagers from Washington, DC............................................................ 196
Table 9.5.2. Median pause duration and median speech rate for interviewee and for
Carissa, the interviewer, for each Washington, DC interview................................. 200
Table 10.3.1. Temporal sequencing variables ............................................................ 212
Table 10.4.1. Petty Harbour (ð) use (from Van Herk et al. forthcoming)...................... 215
Table 10.4.2. Subset of Petty Harbour data................................................................. 217
Table 10.4.3. % stopped and Ns for (ð)..................................................................... 219
Table 10.4.4. Regression analysis for (ð) for all speakers.............................................. 220
Table 10.4.5. Regression analysis for (ð) for young females......................................... 222
Table 10.4.6. % fronted and Ns for (ing)..................................................................... 224
Table 10.4.7. Regression analysis for (ing) for all speakers.......................................... 225
Table 10.4.8. Regression analysis for (ing) for young females..................................... 227
List of Figures

Figure 1.4.1. A schematic indicating the levels of data and metadata for spoken language recordings as used in sociolinguistic research (from Kendall 2008a) .................................................. 18

Figure 2.2.1. Assorted screenshots from SLAAP, discussed in the text (from Kendall 2007a) .......................................................... 30

Figure 2.2.2. SLAAP screenshot of speaker record ................................................................. 31

Figure 2.2.3. SLAAP screenshot of project overview page .................................................. 32

Figure 2.3.1. Four presentations available in SLAAP of the same transcript data (from Kendall 2007a) ........................................................................................................................................... 35

Figure 2.3.2. SLAAP screenshot showing a transcript line with phonetic data.............. 38

Figure 2.3.3. SLAAP screenshot of a tabulation form and audio player......................... 39

Figure 2.3.4. SLAAP screenshot of variable tabulation summary ...................................... 41

Figure 2.4.1. Re-conceived model for abstraction in sociolinguistic data (from Kendall 2008a) ........................................................................................................................................... 43

Figure 3.3.1. Screenshot of Praat editor window ................................................................. 62

Figure 3.4.1. Crossover effect for postvocalic $r$ in New York City (from Labov 1972: 114) ........................................................................................................................................... 67

Figure 3.4.2. Spectrogram with text and pitch displayed .................................................. 68

Figure 3.4.3. Variation score for “Heath” (from Podesva 2008) ............................................ 69

Figure 3.4.4. A Henderson graph example (from Levelt 1989: 127).................................. 70

Figure 3.4.5. Graphicalization screenshot from SLAAP ..................................................... 72

Figure 3.4.6. Henderson graph screenshot from SLAAP ..................................................... 73

Figure 4.2.1. All speakers plotted by year of birth ........................................................... 83

Figure 4.3.1. SLAAP screenshot of transcript summary page ........................................... 85

Figure 4.3.2. SLAAP screenshot showing summary statistics for a transcript ............... 86

Figure 5.3.1. Screenshot from SLAAP showing an analysis of pause ......................... 102
Figure 5.4.1. Histogram of log-pause durations.................................................................109
Figure 5.4.2. Pause duration by speaker...........................................................................110
Figure 5.4.3. Boxplot showing pause distribution for African Americans by region.....117
Figure 5.5.1. Pause duration by pause type and adjacency to filled pause......................122
Figure 5.5.2. Pause durations by pause type and filled pause, by region.......................123
Figure 6.4.1. Screenshot from SLAAP showing speech rate analysis feature..............138
Figure 6.5.1. Distribution of speech rate measurements ..................................................140
Figure 6.5.2. Speech rate by speaker, ordered by median speech rate...........................141
Figure 6.5.3. Speech rates by ethnicity and gender.........................................................146
Figure 6.5.4. Speech rate as a function of utterance length..............................................150
Figure 6.5.5. Number of tokens for each utterance length.............................................151
Figure 7.2.1. Speech rate and pause by ethnicity and gender, with best-fit lines..........157
Figure 7.2.2. Speech rate and pause plots for six ethnicity and region pairs...............161
Figure 8.2.1. Graphic comparison of selected variables from Table 8.2.1.....................170
Figure 8.2.2. Graphic comparison of variables from Table 8.2.2.................................172
Figure 8.2.3. Graphic comparison of variables from Table 8.2.3.....................................173
Figure 8.2.4. Changes in speech rate and pause duration for three speakers ..........174
Figure 9.3.1. Boxplots of effect of interviewer and participant gender on speech rate ..183
Figure 9.3.2. Boxplots of effect of interviewer and participant gender on pause.........185
Figure 9.3.3. Boxplots of the influence on speech rate of different/same ethnicity of interviewers and interviewees .................................................................186
Figure 9.3.4. Boxplots of speech rate based on participant ethnicity and the ethnicity of their interviewer(s)......................................................................................187
Figure 9.3.5. Boxplots of number of participants on speech rate and pause ..........189
Figure 9.4.1. Speech rate and pause duration for EH and her interviewers ..........193
Figure 9.5.1. Distributions of speech rate and pause duration data .......................... 197
Figure 9.5.2. Plot of speech rate and pause duration for DC interviewees ............... 198
Figure 9.5.3. Pause duration and speech rate comparison for Carissa and DC interviewees ........................................................................................................ 201
Figure 9.5.4. Distributions of speech rate and pause data, including Carissa .......... 202
Figure 10.3.1. Example of a Henderson graph (from Levelt 1989: 127) ............... 211
Figure 10.4.1. Screenshot from SLAAP showing a Henderson graph and a single slope line for speaker d ................................................................. 218
Figure 10.4.2. Favorability of stopped (ð) according to ΔSLOPE........................... 223
Acknowledgements

The Sociolinguistic Archive and Analysis Project has been made possible with support from a number of sources. Firstly, I thank the North Carolina State University Libraries for funding the initial development of the project and for the ongoing hosting of the web server and technical support. A number of people at the Libraries have been crucial for the success of the project, including Susan Nutter, Carolyn Argentati, Kristin Antelmann, Amanda French, Troy Hurteau, and the members of the Digital Library Initiatives. SLAAP’s successes in many ways are evidence of the benefits to social science and humanities research that come from a strong relationship with libraries and library scientists.

I am also extremely grateful to the William C. Friday Endowment at NCSU for funding aspects of the project and to Walt Wolfram for many cups of coffee (not all of which were spilled on himself or me). I also thank the Duke University Graduate School for a Summer Research Fellowship, which supported my work on SLAAP in the summer of 2007, and a Dissertation Fellowship, which supported much of the writing of this dissertation.

The data contained in SLAAP and used for Part 2 of this dissertation were collected in projects funded by the National Science Foundation grants BCS-0236838, BCS-0542139, BCS-9910224, SBR-9319577, and SBR-9616331. I thank the NSF for their continued support of the advancement of linguistic science.
The data contained within SLAAP, and analyzed within this dissertation, have been collected, digitized, annotated, and transcribed by a vast number of researchers. I cannot possibly thank all of the people who have contributed data or transcripts to SLAAP here by name, but I would like to thank the following researchers (in no particular order), all of whom contributed recordings that have been used in the analyses of this dissertation: Walt Wolfram, Erik Thomas, Natalie Schilling-Estes, Kirk Hazen, Clare Dannenberg, James Peterson Jr., Drew Grimes, Ryan Rowe, Sarah Hilliard, Jeannine Carpenter, Phillip Carter, Erin Callahan, Mary Kohn, Danica Cullinan, Christine Mallinson, Becky Childs, Gerard Van Herk, and Jennifer Thorburn (I have surely forgotten to mention a few people – I am sorry!). Finally, I am especially grateful to Erik Thomas, who has actively contributed more recordings, more transcripts, and more annotations to SLAAP than any of my other colleagues. Danica Cullinan and Leah White have also been crucially helpful in organizing the digitization of many of the NCLLP recordings over the past couple of years.

Further, more people have been inspirational and/or helpful for both SLAAP and this dissertation than I can possibly list here. I make just a few exceptions to my decision to delay thanks to these colleagues: Daniel Ezra Johnson and David Neal have been wonderfully helpful in answering questions regarding the statistical analyses of the data in Part 2 (of course, any errors are my own). Additionally, I must thank Charlotte Vaughn, who has been influential on all my work, and on SLAAP in more ways than is likely obvious to SLAAP’s users or readers of this dissertation. It is surely clear to anyone who knows the two of us, however, that all of my graphics (here and in recent
publications and posters) that don’t look terrible are her doing. I am grateful to have had Charlotte’s partnership for so many things. Most recently, I thank her for comments on various parts of this dissertation.

(I don’t thank my family in my acknowledgements because they know that I love them.)

Finally – and even though I believe (or at least hope) that they might pass me without these words – I thank my dissertation committee Walt Wolfram, Agnes Bolonyai, Ron Butters, and Erik Thomas for being more than great mentors. I am so glad that I have been able to work with you all, and call you my friends.

Many sections of this dissertation draw on papers I have previously published or presented. In particular, much of Chapters 1 & 2 reprint Kendall (2008a), Chapter 8 expands on the analysis in Kendall and Wolfram (forthcoming), and §9.5 presents data discussed in Mallinson and Kendall (forthcoming). The analysis of §9.5 has also been discussed in conference papers (Mallinson 2007; Mallinson and Kendall 2008; Kendall, Mallinson, and Whitehead 2007). Chapter 5 continues work presented in Kendall (2006, 2007c). Finally, Chapter 10 was presented in Kendall (2008b). Drawing on recordings collected by others, of course, means that I also draw on the research conducted by others. I try to acknowledge specific instances of this in the text where appropriate.
1. Introduction

1.1. Rationale and overview of the present work

Linguists’ data come in many forms. From grammaticality judgments to reaction times to acoustic measurements, linguists build their theories and understanding of language through a wide variety of data types. Though the notion of what constitutes primary data can differ from research project to research project and from researcher to researcher, within sociolinguistics – the study of language in its social context – “data” often involve some sort of empirical language recordings, such as recordings of naturally occurring speech.

The reliance on naturalistic spoken data is so profound in sociolinguistics that a large proportion of energy spent developing sociolinguistic practices has focused on the refinement of the sociolinguistic interview as a method for the acquisition of naturalistic, conversational speech (cf. Labov 1984). Despite this extensive focus on interviewing strategies, however, with few exceptions (e.g. Poplack 1989; Tagliamonte 2006; Kendall

---

1 But, see more recent work, such as Macaulay 1991, Schilling-Estes 2004, and Bucholtz 2006, which examine the “interview” more critically.
sociolinguists have not similarly addressed the management (e.g. organization, storage, accessibility, and preservation) of their data once they acquire it, or how to ensure future access to it. For example, there have been few discussions of strategies for how to best organize a project’s recordings and label the project’s interviews, tapes, and speakers (but see Tagliamonte 2006), or for how to best code linguistic features (but see Wolfram 1993, 2006, and Blake 1997). As a result, spoken language data management practices have remained individualistic within sociolinguistics, despite work in other linguistic subfields (such as in documentary linguistics, cf. Himmelmann 1998; Bird and Simons 2003) to standardize and improve these practices. That is, each sociolinguist typically has his or her own set of practices for dealing with data and there are no conventions for how (or whether!) he or she should report on how the data have been treated through all the steps from data acquisition through analysis and from interpretation to presentation.

This state poses a number of relatively unexplored problems – and opportunities – for the field. How reliable and replicable are our findings? How useful and accessible are our data for addressing new questions and re-thinking old ones? That is, can the data collected for project X be used to shed light on research question Y? Are our annotations, notes, transcripts, and data tabulations in formats that are maximally useful or even comparable to one another? These questions, and many more, are rarely addressed in the sociolinguistic literature yet have far reaching implications for our work.

This dissertation arises from my work over the past few years to address these problems and opportunities through the development of the Sociolinguistic Archive and
Analysis Project (SLAAP), a fairly large-scale attempt to build a sophisticated sociolinguistic data management system. The bulk of the present work – Part 2 of this dissertation – concerns a number of explorative case studies that leverage the data in SLAAP to examine variation in speech rate and pause and the relationship between these speech features and more “normal” sociolinguistic variables. First though, in Part 1, I explicate SLAAP, its data model, implementation, and features, and, following Kendall (2008a), I attempt to make some headway towards improving the explicitness and robustness of sociolinguistic data management strategies more generally.

The remainder of this chapter briefly reviews the history of data management practices within sociolinguistics and some related empirical linguistic disciplines, insofar as these practices have been discussed within the literature. It highlights advances that have been made in other subfields of linguistics, such as documentary linguistics and corpus linguistics, and argues that these advances have not sufficiently been synthesized into common sociolinguistic practice.

This sets the stage for Chapter 2’s overview of the Sociolinguistic Archive and Analysis Project (SLAAP), the web-based digitization and preservation project being carried out by the North Carolina Language and Life Project at North Carolina State University. SLAAP aims to highlight new directions for sociolinguistic data management practices as well as the analytical and theoretical benefits to which new approaches to data management can lead (Kendall 2005, 2006-2007, 2007a, 2008a; Kendall and French 2006; SLAAP User Guide 2008).
Chapter 3 discusses the transcription and data model that drives the SLAAP software. In particular, it discusses the complex theoretical and methodological issue of what counts as a unit of speech data, arguing that a definition of the utterance on purely acoustic grounds—as speech separated by measurable periods of silence on the part of the speaker—provides a powerful *methodological* unit for analysis. Chapter 3 also contains discussions of SLAAP’s transcription conventions as well as visualization techniques for speech data, using examples and features from SLAAP to demonstrate the utility of alternative visualizations for the analysis and understanding of speech data.²

Chapter 4 introduces the reader to Part 2, interconnected “explorations” of the data housed in SLAAP. Part 2—the larger part of this dissertation—could be read as examples of some ways that SLAAP’s data model can enhance sociolinguistic analysis, but it is primarily meant to substantively contribute to an understanding of variation in pause and speech rate, and the relationship between this variation and more “normal” sociolinguistic variation (that is, morphosyntactic and phonological variation; variables of the sort described in Wolfram 1993). Chapters 5 and 6 then seek to extend the sociolinguistic lens to quantitative analyses of pause duration and speech rate respectively, pursuits often considered outside the realm of sociolinguistics (cf. Macaulay 2002). Specifically, these chapters conduct large-scale, corpus-like studies to assess

² Part 1 of this dissertation highlights the *conceptual* and *analytical* benefits of rigorous data management strategies. While it touches on the *technical* benefits of following “best practices” in the management of linguistic data, it does not purport to provide a thorough discussion along those lines. For example, I do not provide a technical discussion of the digitization process and I only briefly touch on the mechanics of time-aligned transcription. The *SLAAP User Guide* (2008) touches on some of these issues, but see Bird and Simons (2003) or the OLAC website (http://www.language-archives.org/) for good discussions of “best practice” guidelines for language resources.
whether these speech features can be seen to correlate with social characteristics (e.g.,
gender, ethnicity, etc.) of speakers.³

Chapter 7 builds on the independent analyses of pause and speech rate by looking
at the relationship between the two features, which we might intuit to be related in one
way or another. Chapter 8 continues this exploration by presenting case studies of a
handful of speakers to examine speech rate and pause more closely, in terms of stylistic
and register variation. Chapter 9 pursues a related venture and re-considers pause and
speech rate as “interactional variables” through three different approaches. It asks such
questions as what interactional and conversational factors impact pause realization and
speech rate. Chapter 10 considers how the examination of pause and speech rate informs
our understanding of “traditional” sociolinguistic variables by proposing and evaluating a
combined metric of “sequential temporal patterning”. In so doing, it returns to more
traditional considerations of pause and speech rate to examine their role specifically as
factors in the realization of other variables and more generally as paralinguistic cues to
speech style. It is commonsensical that speech rate, for example, might have an impact
on variable realizations (whether some sort of phonetic or phonological reduction or even
triggering a morphosyntactic alternation). However, through SLAAP, this relationship
can be systematically and quantitatively analyzed, bringing empirical evidence to bear on
long-standing assumptions and questions. Chapter 11, finally, seeks to provide closure

³ When I began working on pause and speech rate, studies of pause and speech rate seemed mostly in the
memory of psycholinguistics (e.g., Rochester 1973) and not in current sociolinguistic vogue. Admittedly,
this seems to be changing and some of my original questions (e.g., Kendall 2006, 2007c) of whether pause
and speech rate are worth examining from a language variation perspective are increasingly unnecessary.
See, for example, not only my own work but also recent papers like Clopper and Smiljanic (2007) and
Salmons, Jacewicz, and Fox (2008).
by summing up the work presented here and highlighting areas for future work, both in substantive terms and for the Sociolinguistic Archive and Analysis Project in general.

1.2. A short history of (the treatment of) data within sociolinguistics

Since the foundational work of William Labov (1963, 1966) established the examination of linguistic variation and its social correlates as an important approach to understanding language, a number of published accounts have described sociolinguistic methods and practices. In the early years of sociolinguistics, most major projects (e.g. Labov 1966; Wolfram 1969; Sankoff and Sankoff 1973; Trudgill 1974; etc.) published thorough accounts of their methods, ranging from explications of their sampling techniques – how and why they chose the informants they did – to discussions of their interviewing techniques and even of training their fieldworkers (cf. Shuy, Wolfram, and Riley 1968). These methodological reports were an important – and probably necessary – step in establishing sociolinguistics as a credible and quantitatively oriented social science. These publications also served to aid future scholars by explicitly sharing “the knowledge of the problems [the researchers] faced and the solutions [they] tried” (Sankoff and Sankoff 1973: 12).

In 1974, Wolfram and Fasold’s publication of The Study of Social Dialects in American English began a tradition of leading scholars publishing textbooks on sociolinguistic methods (cf. Milroy 1987; Milroy and Gordon 2003; Tagliamonte 2006). These textbooks have typically been widely read by practitioners in the field. Used not
only by students but also by colleagues in the design of their research projects and fieldwork, they have been influential in the development of shared methodological practice within sociolinguistics.

However, throughout both these research reports and methodological textbooks, authors’ discussions of methodology almost always jump from data acquisition to data analysis and/or to demographic and theoretical issues pertaining to analyzing language in relation to social attributes of speakers (as in Wolfram and Fasold 1974; Milroy 1987; Milroy and Gordon 2003). There are numerous robust discussions of issues like how to choose informants and how to elicit and obtain “good” speech, but, throughout the literature, there are rarely discussions of the many intermediate steps and decisions that occur between having some speech recorded on a device and sifting that speech down to “data” for analysis. In sum, while there have been – and continue to be – discussions in the literature on approaches to sociolinguistic methodology, almost across the board these have neglected to discuss issues in data management.

Of course there are exceptions. As mentioned above, many of the foundational early sociolinguistic projects have entire publications dedicated to their field and lab methods, and a handful of these do at least touch on aspects of their data management.

4 The discussions about recording “good” speech, however, center on methods to elicit and record naturalistic, or casual, speech. None of the major overviews of sociolinguistic field methodology give more than cursory mentions of the technical aspects of recording interviews – which kind of recording device to use, what kind and how many microphones to use, and where to place them, etc. This, I believe, remains a major methodological problem for sociolinguistics, but addressing this is, unfortunately, outside the scope of this dissertation.

5 Wolfram (1993) provides perhaps the best published account of how one can move from a speech recording to a quantitative set of data for analysis by elucidating the heuristic of the sociolinguistic variable. Yet, it too takes as its starting point a collection of recordings ready for analysis and as its implicit end point a set of quantified “data,” complete and ready for assessing in terms of social correlates.
For example, Labov (1984: 52) ends with a short section mentioning the Philadelphia Language Change and Variation Project’s archive of recordings and describing its size and who can access the archive. Shuy, Wolfram, and Riley’s (1968) book provides a more thorough account of the treatment of the Detroit Dialect Study’s data and recordings than most other projects give; it provides a rather full accounting of their field methods, explaining and commenting upon everything from determining a sampling system to computer coding their data and from hiring interviewers and an administrative assistant to training them in field interviewing and phonetic transcription. Similarly, Sankoff and Sankoff (1973) thoroughly outline the field and laboratory methods for their sociolinguistic study of Montréal French, including a discussion of their computerized transcription system and an overview of their complete database, which they enumerate in detail:

i. 120 reels of taped interviews (2 copies);
ii. 64 boxes, most of them full, of computer cards containing transcriptions, about 100,000 cards in all;
iii. computer printouts (in several copies) in readable format;
iv. in addition, we are presently storing corrected transcriptions on a master computer tape. Thus, to date, 40 interviews, over 20 boxes of cards, are now stored on a single reel of tape at the Centre de Calcul (Sankoff and Sankoff 1973: 42).

At first glance, it may seem that the sort of detail provided by Sankoff and Sankoff (1973) is superfluous. That is, one might ask: what benefit do readers gain by their listing the detailed contents of their linguistic closet? Yet, I argue, even the simple list of the data provided by Sankoff and Sankoff (1973) is greatly useful. On one level it provides readers with specific and straightforward information about the data collection
(how it is organized and where one can find it). On another level it indicates the ways in which the analysts interacted with their data.

This is not to say that discussions of data and their treatment have been altogether missing from more recent sociolinguistic literature. Poplack (1989) provides an excellent discussion of the Ottawa-Hull French Project’s data archive and methodology, a project with a goal to improve methodologies inherent in working with large sets of data for sociolinguistic analysis. She points out that one area in which development has been sporadic at best is in the construction of major sociolinguistic data bases. The trade-off between sociological representativeness and ethnographic thoroughness has resulted in insufficient data from a large sample of speakers, or masses of data of questionable generalizability from a few speakers. Efforts to increase quantity or authenticity of recordings are still marked by losses in the quality of the data obtained. And even as a data base reaches respectable size, its accessibility is concurrently hampered by the uneconomical effort needed to search it systematically in studies of individual variables (Poplack 1989: 413).

Her paper further provides the most thorough discussion in the literature of many of the steps from the determination of a sample population, to interviewing and recording that population, to organizing the resulting collection, and developing a computer-based corpus of the recordings. In addition to the excellent scope of Poplack’s (1989) publication, it should be noted that the Ottawa-Hull French Project itself is exceptional in its size, containing approximately 270 hours of speech and over 3.5 million transcribed words (Poplack 1989: 429).

Returning to textbooks, Tagliamonte’s (2006) Analysing Sociolinguistic Variation, the most recent textbook on sociolinguistic practice at the time of this writing, may be an indication that data management strategies are becoming more explicit within
sociolinguistics itself. Her text has an entire chapter, “Data, data, and more data,” that reviews a wide range of data management tasks, from labeling and organizing interviews into a coherent corpus to transcribing the data and working with computerized transcripts and recordings.

Despite Tagliamonte’s (2006) recent intervention, which is hoped to be an indication of a growing awareness of the importance of explicit documentation of data management methods, many sociolinguists are still not particularly good at preserving and managing their often large collections of data as a result of this general paucity of focus over the years. Furthermore, I would argue that a potentially unnecessarily large portion of the sociolinguistic enterprise is spent on redundant data collection (i.e. conducting new interviews when there might be data near-at-hand that could speak to the present research question) and (re-)analysis, since existing data collections are frequently not well-organized or accessible for future work.

1.3. How does speech constitute data?

One of the primary issues that arises when considering the management of data is the fact that the term data itself means a lot of different things to a lot of different sociolinguists (and more broadly, of course, linguists). That is, throughout many of the diverse approaches subsumed under the rubric of sociolinguistics the term data is used to
refer loosely to some sort of captured real-world speech event, but the data on which analyses are based differs greatly depending on the theoretical perspective of the researcher and the substantive questions being asked. At the level of analysis, variationist sociolinguists, for example, work with a very different sort of data than do, say, discourse analysts or linguistic anthropologists.

To take a specific example, let’s consider the quotative system in English. Since Butters (1982) mentioned the use of *be like* as a preface to a quotation (e.g. “And I’m like ‘No way, man!’”), a number of studies (e.g. Romaine and Lange 1991; Ferrara and Bell 1995; Tagliamonte and Hudson 1999; Buchstaller 2006; Kohn and Askin 2008) have investigated the increasing use of *be like* as a quotation introduction. While many studies of quotative *be like* (such as those listed above) are conducted in similar ways – roughly by counting instances of *be like* and other quotative markers (such as *say* or *go*) and comparing the relative occurrences of each with other linguistic features of the discourse and social attributes of the speakers – a study of the feature could just as well be undertaken in widely different ways. Different approaches could depend not only on the researcher’s theoretical background, but also on the format and type of data used for the analysis (or even on the original data sample or instance that lead the researcher to be interested in the particular analysis in the first place). So, a researcher who becomes interested in studying *be like* from seeing its use in transcripts may conceptualize a study very differently – most likely focusing on morphosyntactic, discourse, or corpus-based

---

6 Of course, there are a number of sociolinguistic pursuits that do not focus on speech recordings as their primary data (e.g. perceptual dialectology, cf. Preston 1989). As is clear to the reader, those sorts of data are not the focus of the present discussion.
approaches – than one who hears instances of be like in audio recordings, where phonological and phonetic aspects of the occurrences – like variation in intonation and stop-release, or the impersonation of others’ “voices” – may be more striking.

In other words, two analysts working within the same theoretical framework, even potentially on the same set of substantive issues, may have very different conceptions of what constitutes the data for their analyses. They may also have different views on just what exactly their substantive issues are depending on the nature of the data they have at their disposal. The different views or approaches in turn can have dramatic effects on the researchers’ findings and can affect whether or not those findings can (or should) even be compared to one another.

In short, different ways of conceptualizing speech as data have different outcomes, whether or not those outcomes are intentional. To further exemplify this problem, let us turn briefly to two primary notions of data used within sociolinguistic disciplines: variable tabulations and orthographic transcriptions.

1.3.1. Variable tabulation in practice and theory

Tagliamonte tells us “the advantage of variation analysis is working with real data, often from representative samples of communities, and from scrutiny of hundreds and perhaps thousands of instances of the linguistic variable” (2006: 74-5). To isolate those “hundreds and perhaps thousands” of variable realizations involves a great deal of work on the part of analysts as they filter the complexities of speech down to analyzable
formats. Variable tabulating – the extraction and coding of different realizations of the same linguistic variable – is the central methodology within quantitative variationist sociolinguistics and the primary means for undertaking this filtration process (cf. Labov 1966; Wolfram 1993, 2006).

While sociolinguists may agree in principle on how they study language variation, there is no agreed upon set of standards for how they move from a real-world speech event to a set of quantified data about a particular linguistic feature. Wolfram (1993) gives some guidelines for determining and evaluating linguistic variables, but also points out a number of problems with the heuristic – such as the difficulty (and under-reporting within the literature) of how to determine what range of variation is subsumed by a single *linguistic variable* (see also Wolfram 2006), as well as problems of inter-analyst agreement and even intra-analyst reliability.

Blake (1997) and Rickford, Ball, Blake, Jackson, and Martin (1999) provide excellent (and exceedingly rare) discussions of some of these issues. Blake (1997) investigates the forms not counted by different groups of researchers in the study of copula deletion in African American English, while Rickford et al. (1999) examine how frequencies of copula contraction and deletion have been computed by different groups of researchers and how the differences in methodology affect the results of the analysis. Rickford and his colleagues, in particular, demonstrate that different theoretical assumptions and views about data impact the quantitative outcome of a study with consequences of greater theoretical and descriptive import. The significance of these
papers is punctuated by the fact that there are countless variable tabulation issues that remain hidden, undiscussed in the literature.

Finally, it must be noted that this sort of approach on its own has larger potential problems, as Macaulay, for example, notes in a discussion of the sociolinguistic interview.

Interviews as a whole have not been used as data. It is somewhat paradoxical that most of the speech collected in sociolinguistic surveys remains unanalyzed. Most investigators have followed Labov’s lead in concentrating on a few variables. In this approach a certain number of tokens are extracted from the interview and coded. The analysis then deals with these tokens and the remainder of the interview is ignored… The concentration on such variables has, however, an influence on the kind of questions that are asked (1991: 5).

In other words, the focus on a handful of quantified linguistic features divorced from their original context ignores a great deal of the data. In addition to limiting the questions that can be asked, it limits the answers that can be obtained. So, while often a necessary procedure, a research focus on just a few variables or linguistic features can often lead to only a partial understanding of the issue being investigated. Thinking further about quotative be like, might morphosyntactic or text-based corpus analyses be missing important aspects of the variation?

1.3.2. Orthographic transcription in practice and theory

Both within and outside linguistics, the orthographic transcript is the primary representation used to present speech in a non-aural format. Within language research, it
is often the chief mediating apparatus between theory and data, yet the act of transcription is often undertaken as a purely methodological activity, as if it were theory-neutral. As Mishler (1991: 261) wrote,

videocameras with microphones have replaced the camera with its lens, and “nature records itself” on magnetic tape. And then, as if we were printing positives from negatives, we inscribe the sounds in writing. Each of these steps of re-presentation is a transformation and each may be made in many different ways.\(^7\)

Further, each potential way that a transcriber makes that “re-presentation” has an effect on and constrains the resulting possible readings and analyses (Ochs 1979; Bucholtz 2000; Edwards 2001). Decisions as seemingly straightforward as how to lay out the text, to those more nuanced – like how much non-verbal information to include and how to encode minutiae such as pause length and utterance overlap – have far-reaching implications on the utility of a transcript and the directions in which the transcript may lead the analysts.

Despite the widely agreed upon stance that “transcriptions are not substitutes for the original recordings but additional tools which can be used to help analyze and understand these recordings” (Liddicoat 2007: 13, on transcripts in conversation analysis; see also Tagliamonte 2007), transcripts all too often become the sole means of exemplifying and sharing access to speech recordings. And it is also possible these non-aural re-presentations often do end up becoming the data used for analysis in lieu of the audio (or video) recordings, despite the frequent warnings to the contrary.

\(^7\) This quote, from Mishler (1991: 261), is important at a number of theoretical levels concerning abstraction in the recording process. Unfortunately, it is outside the scope of this dissertation to fully divert into a discussion along those lines.
Some of the time scholars do explicate their practices regarding their medium of analysis and when they do this it is often shown to be a valid technique. For example, Poplack (1989: 436-7) wrote, “we thus feel confident that the Ottawa-Hull corpus may be used to study morphosyntactic and lexical phenomena without the necessity of recourse to the audio tapes.” Others (e.g. Rickford and Théberge-Rafal 1997; Buchstaller 2006) have also shown that many kinds of morphosyntactic and lexical studies can be carried out successfully on transcripts. However, what is concerning here is that since no standard publication practices exist for explicating analytic methods, it is quite possible that any number of studies do base their analyses on transcripts instead of the source recordings in ways that are problematic.

A major part of the problem behind the use of transcripts for language research is that the text of a transcript is always an incomplete (and interpreted) record of the original interaction (Edwards 2001). Transcripts can be improved, however, through the use of computers (Kendall 2005, 2006-2007, 2007a; MacWhinney 2007), in ways that can mitigate some of these central problems, a point to which we will return in the next chapter.

1.4. Data vs. metadata

But, what are variable tabulations and what are transcripts? Are they really data? In a sense, they often are, but in a better sense they should perhaps be considered
metadata, data about the core data. They are abstractions, representations of the real-world speech events that actually constitute the core data for sociolinguistic research. This statement on its own may not be particularly striking, but it has, I argue, important consequences. Metadata – such as variable tabulations and transcripts – are inseparable from their source data and should not be analyzed as things in their own right, divorced from their source recording or real-world interaction.

Clearly, abstracting from the original speech event is an integral part of analyzing that speech event. Despite our best attempts, we can never analyze the “fleeting events of an interaction” (in the words of Edwards 2001: 321) without using some sort of representation (such as a transcript or a spreadsheet of variable tabulations) as a proxy for the data. The important point is that analysts must always keep in mind their core data – not even the recording itself, but rather the never-fully-reconstructible real-world interaction that underlies the recordings.

Figure 1.4.1 (from Kendall 2008a) provides a schematic illustrating the ways in which speech data are traditionally abstracted in their various representations. Each level in the schematic should be considered a layer of metadata over the real-world speech event, the actual data of a particular dataset. The real-world event is shaded and connected to the recording without arrows to represent the fact that it is evanescent, over as soon as it is begun, and rarely (if ever) available to the researcher for deep analysis.
In Figure 1.4.1, even the connection between the speech event and the recording is represented with a dashed line because the very act of recording is, in itself, an abstraction and therefore does not have a solid connection to the event, the actual *data*. In the schematic, the recording is described as “*data*” (in quotation marks) because it is...
often used in place of the original interaction, as if they were one and the same. A video recording captures more of the real-world interaction than an audio recording, but even then the video recording may be missing important action (e.g. that occurred off camera). It is also still an abstraction in the sense that it is always ultimately divorced from the real-world context(s) in which the original speech event occurred. Ethnographic inquiry can help to strengthen the connection between a recording and its original context, but even then there is only so much an analyst can know about the full scope of an interaction.

Beyond the recording, each further representation in Figure 1.4.1 is a further abstraction, and as such each is illustrated with a dashed connector line, sometimes connecting to the recording, but sometimes also to other layers of metadata, multiply removed from the original speech event. Arrowheads indicate the directions to which some layers of metadata contribute to the accuracy and understanding of other layers. Distance is also used to indicate the degree of abstraction from the core data. Through this depiction we see that many of the quantified data types that are most useful for analysis are often the most abstract.9

This is not to argue that these abstractions are “bad.” They are, without doubt, necessary mediations between a spoken interaction with its full complexity and the categories of meaning of interest to the analyst (cf. Edwards 2001). Variable tabulations

---

9 This conception does “privilege” instrumental quantitative data over impressionistic quantitative data since instrumental data, such as vowel formant measurements, are measured directly from the recording and, therefore, can be considered to be closer to the recording. Of course, a number of factors still contribute to the abstractness of acoustic measurements, such as the facts that analysts determine where in the speech to measure based on interpreted cues and choose the settings to use for the measurements based on impressionistic assessments of the speech.
and other quantified metadata are often highly useful for understanding the social world, in part by virtue of their level of abstraction. In sum, I argue that analysts should keep in mind the distance a particular set of metadata may have from the speech event itself, to be mindful about that distance in each step of the analysis, and to be explicit about it in publication and presentation.

1.5. Treating speech as data in the 21st century

In 1973, Sankoff and Sankoff proposed that, while “detailed descriptive studies of complex urban communities involve a good deal of painstaking and sometimes tedious work[,] much of this can, however, be reduced by maximal use of automated, computerized data processing techniques” (1973: 47). Yet, thirty-five years later, most approaches to sociolinguistic data management have not moved towards computerized techniques and even those that have, remain individualistic, ad hoc, and often implicit. Sankoff and Sankoff’s (1973) work may have been ahead of its time, but we are now well into an era in which the technological advancements of the growing digital age enable vastly improved computerized analysis and storage of speech recordings.

In some ways this is to argue that sociolinguistic practice should move closer to fields such as corpus linguistics and natural language processing. This is neither a bad

---

While I do not discuss the relationship between “the social world” and the “real-world speech event” here, I do intend to hint at the complexity of that relationship in this schematic. To a large degree sociolinguists investigate language as it’s contextualized in the social world. Yet, it is not always entirely clear how a particular speech event relates to the larger “social world” of interest.
thing nor new thing. Sociolinguistics and sociolinguists have a lot to learn from related fields, and the ever improving technologies for language analysis make some convergence inevitable. The use of statistics within sociolinguistics is an excellent example of this. The traditional statistical framework used within sociolinguistics – variable rule analysis (Cedergren and Sankoff 1974) – first through the tool Varbrul (cf. Pintzuk 1987) and its later versions, and more recently through GoldVarb (Sankoff, Tagliamonte, and Smith 2005), enabled decades of important sociolinguistic work. Yet, as Johnson (2009) points out, the traditional sociolinguistic Varbrul-like implementation of logistic regression is limiting and its proprietary style of result output has isolated variationist sociolinguistics from sister disciplines. Johnson’s (2008, 2009) new Rbrul package bridges the gap for sociolinguists by expanding traditional variable rule analyses to include continuous predictors and responses and random effects (i.e., accounting for the role of individual speakers). Rbrul also presents its results in standard formats, matching more commonly used conventions in other fields. In many ways, the movement by sociolinguists towards more broadly used statistical methods also leads, I believe, towards greater convergence with other linguistic and social scientific disciplines.11

Meanwhile, disciplines outside of sociolinguistics are increasingly focusing on spoken language data management practices through computer-based methods.

Associations like the Open Language Archives Community (OLAC; cf. Bird and Simons 11

11 In the analyses of Part 2, I make extensive use of Rbrul (Johnson 2008, forthcoming) for statistical analysis. While these analyses are framed and described as sociolinguistic studies, they draw on, and in many ways are comparable to, work that might be best described as corpus-based, computational, psycholinguistic, and so forth.
2003) and E-MELD (cf. Aristar-Dry 2002) are working towards the development of inter-operable and standards-based archives for language documentation. Bird and Simons (2003), for example, provide a thorough enumeration of “best practice” guidelines at the same time as they highlight often overlooked potential problems that surround building comprehensive language archives. These efforts appear to be finding success and ought to be informing sociolinguistic practice.

Corpus linguistics has grown dramatically in recent years as a both a subfield of linguistics in its own right but also as a set of methods used throughout the linguistic disciplines, with tools and corpora becoming more developed, available, and useful for a diverse range of pursuits (McEnery and Wilson 2001). With the growth of the Internet and the rise in corpus and computational approaches to linguistic analysis, more projects across all empirical linguistic disciplines are being explicitly designed around the generation of coherent data collections and more publications are appearing on data collections. Even academic libraries are becoming involved in the development and publication of spoken language collections (Kendall and French 2006; Cooper 2007). Groups such as the Linguistic Data Consortium (LDC) at the University of Pennsylvania (cf. Cieri and Liberman 2006), the TalkBank project (cf. MacWhinney 2007), and the Oxford Text Archive have developed and made available a wide range of language corpora and corpus analysis tools.

Many of these corpus-generation projects are of direct interest to sociolinguists (e.g. Pitt, Johnson, Hume, Kiesling, and Raymond 2005; Kretzschmar, Anderson, Beal, Corrigan, Opas-Hänninen, and Plichta 2006; Kendall 2007a; see also Bauer 2002). Beal,
Corrigan, and Moisl’s recent two-volume set of texts (2007b, 2007c), *Creating and Digitizing Language Corpora*, contains papers discussing a range of “unconventional” corpora – corpora containing sociolinguistic variation – and, as such, has a great deal to offer to sociolinguists. At the same time, however, many corpora, tools, and other products have not been of great use for sociolinguistic pursuits (as discussed by Kretzschmar et al. 2006: 173-4). In short, Beal et al. are correct to point out that their “volumes are unique, since public output to date has primarily concentrated on describing and assessing the models and methods which underpin conventional corpora and the annotation standards/analytical tools developed specifically for them” (2007a: 2).

The point here is that “normal” sociolinguistic practice could be greatly improved by embracing the sorts of “data collection” mentalities held by documentary and corpus linguists. In fact, the data collection work that sociolinguists undertake as part of “normal” sociolinguistic work should really be viewed no differently than the work that goes into the development of a sociolinguistically relevant corpus (in Beal et al.’s (2007a) terms, an “unconventional” corpus), with the exception, perhaps, of how public the final collection might be.

Poplack’s forward to Beal et al. (2007b) points out that “the projected use of the corpus, as *end-product* or *tool*, is clearly the determining factor” (Poplack 2007: xi, emphasis in original) for how a data collection gets treated by its creators. Many explicit corpus creation projects focus on the construction of an *end-product*, whereas for most sociolinguists the utility of the corpus is its role as a *tool* for researching a particular question. As such, the two groups see their aims as being different and, consequently, the
focus in the corpus linguistics literature – and to a certain degree in the documentary linguistics literature – on compiling corpora has had little traction for sociolinguists. This, at least in part, accounts for the fact that sociolinguistic data collection is rarely treated explicitly as corpus-generation work. Returning to Poplack’s preface, considering sociolinguistic data collection and management as corpus work can be of benefit to sociolinguists because of

the opportunity it affords to serendipitously discover what one wasn’t looking for, to characterize the patterned nature of linguistic heterogeneity, and in particular the hidden, unsuspected or ‘irrational’ constraints that are simply inaccessible to introspection or casual perusal (Poplack 2007: xii).

This is not to argue that sociolinguistics, and all sociolinguists, should move toward corpus linguistic approaches to their data and analyses. In fact, conversely, I would suggest that corpus linguists look to the data collection work of sociolinguists – especially those whose fieldwork is ethnographically informed – for comprehensive, thorough collections of data for their corpora. Nonetheless, there is no doubt that sociolinguists can learn a lot from thinking about their data as corpora, both for their projects at hand and for future – possibly not yet conceived of – investigations.

1.6. In closing: Archiving, organizing, and managing speech data for sociolinguistic analysis

As discussed earlier, a long-standing problem within sociolinguistics is that, despite the concentration on the acquisition of “good,” naturalistic spoken language data, very little emphasis has been placed on how these recordings can be best organized and
how (or whether) they can be preserved for future use, either to re-test the findings of the original analysis or to investigate new research questions. Explicit data management work leads to a better-preserved data collection. Reel-to-reel tapes, cassette tapes, even compact discs and computer hard drives deteriorate and can break or be lost and should be digitized, organized, and backed up.\textsuperscript{12} Since every collection of sociolinguistic recordings is a unique and irreplaceable window into a speech community, the holders of a collection should do their utmost to ensure that it remains complete and accessible, whether it be only for their own use or to share with other researchers.\textsuperscript{13}

Most sociolinguists recognize and take this level of responsibility toward their data. A good archive, however, does more than just preserve audio or video recordings. It creates a level of organization that is more complete and useful than otherwise. It also allows researchers to better back up their claims, with easier access to their data and better means to support their analyses and investigate problematic cases. In short, explicit data management work makes for better analyses.

The next chapter introduces the Sociolinguistic Archive and Analysis Project and seeks to highlight some of the conceptual and theoretical benefits that are obtained

\textsuperscript{12} For information about digitizing audio recordings and some “best practices” advice, see Plichta and Kornbluh (2002) or the National Initiative for a Networked Cultural Heritage’s \textit{Guide} (NINCH 2003). Of course, digitizing, in itself, does not ensure the preservation of a recording. As Bird and Simons (2003) argue, software versions, file formats, and hardware typically have a lifespan of between a few years and a decade. Digitized materials must be actively maintained and/or stored in open, non-proprietary formats and adhere to widely accepted standards.

\textsuperscript{13} There are, of course, numerous ethical and IRB issues concerned with the preservation and, especially, sharing of interview data. Each project may have its own constraints and parameters with regards to how the recordings can be used and how much, or whether, they can be shared with other researchers. Readers are urged to check with their IRB and their contracts with their informants about the limitations on the use and sharing of their data. Access considerations such as these are discussed in the \textit{SLAAP User Guide} (2008).
through a stronger focus on data management and a reconsideration of speech as “data.”

Finally, as Part 2 will explore, these sorts of reconsiderations can also create opportunities to evaluate new research questions.
2. Extending the utility of the archive

2.1. The Sociolinguistic Archive and Analysis Project

This chapter more fully introduces the Sociolinguistic Archive and Analysis Project (SLAAP; http://ncslaap.lib.ncsu.edu/), an exploration of new approaches to storing, managing, and interacting with natural speech data. The project centers upon the creation of an online archive and analytic toolset for the sociolinguistic data collection of the North Carolina Language and Life Project (NCLLP), a research initiative at North Carolina State University.¹ The NCLLP’s large and growing collection of interviews is an important resource for linguists in general and for other scholars interested in the American South. As a part of the SLAAP initiative, all of these sociolinguistic interviews are being digitized. SLAAP also seeks to provide a central repository for sociolinguistic recordings from outside the NCLLP and increasingly is storing larger collections of non-NCLLP materials.

To a certain degree SLAAP looks like some of the other corpus development projects discussed in the literature (such as the ONZE Corpus discussed by Gordon, Maclagan, and Hay 2007). However, SLAAP seeks to fill a gap in terms of

---

¹ The NCLLP is a sociolinguistic research initiative at North Carolina State University with one of the largest audio collections of sociolinguistic data on Southern American English in the world. The growing collection contains as many as 2,000 interviews conducted from the late 1960s up to the time of this writing, most on analog cassette tape, but some in formats ranging from reel-to-reel tape to digital video. For more information about the NCLLP visit the project’s website at http://www.ncsu.edu/linguistics/ncllp/.
sociolinguistic practice more than it seeks to create a particular corpus. In terms of Poplack’s (2007: xi) explanation of corpora design as oriented towards either end-product or tool, SLAAP is very much conceived of as a tool. It is a speech data management system (SDMS), which houses an expanding collection of audio recordings, designed to improve quantitative and qualitative sociolinguistic analysis.

The specific goals behind SLAAP are multiple. At a practical level, it seeks to digitize and preserve a large collection of interviews. It also aims to provide researchers with better access to and interfaces for their data. At a theoretical level, SLAAP questions and rethinks current linguistic and sociolinguistic conceptions of the nature of speech data, its representations, and the sorts of questions that can be asked of it. In short, SLAAP seeks to address some of the longstanding oversights in sociolinguistic methodology addressed in the previous chapter and to evaluate the benefits of a more rigorous, explicit approach to sociolinguistic data management.

2.2. Features of the SLAAP software

The SLAAP software is entirely web-based. As such, its users can access the entire archived collection and all of the speech data management system’s features over the Internet. A number of the software’s features are presented in a collage of

---

2 The archive is actively growing as part of our ongoing digitization and transcription effort. As of December 2008, the SLAAP digital archive contains over 1,050 interviews and over 815 hours of audio. About 31 hours have associated time-aligned transcripts, making a transcript corpus of over 300,000 words. SLAAP currently only houses audio materials, although the software has been designed to be extensible in this regard, and it is hoped that video recordings from the NCLLP collection and elsewhere can be incorporated into the archive in the near future.
screenshots in Figure 2.2.1 (from Kendall 2007a). I now briefly survey some of the main features – that is, the main web pages – of SLAAP in order to exemplify the benefits of this sort of online archive and approach to sociolinguistic data management.

SLAAP’s basic features include: (1) & (2), a browsable and searchable interface to the archive collection, (3) an audio player with an annotation tool that allows users to associate searchable notes to specific times within the audio files (and to listen to those particular passages at the click of the mouse), and (4) an audio extraction feature that enables users to download excerpts of the audio files without having to download or locally store the large files.

The full record view, shown in (2) of Figure 2.2.1, gives users access to the full set of data and metadata available for each interview which they have authorized access to, including information about each speaker and interviewer, when and where the interview took place, notes about the interview, and so on. This view also provides information and metadata about the media files for a particular interview. Some of these data, such as the length of each media file and an approximation of its signal-to-noise ratio, are computed automatically by the software, while other parts, e.g., metadata about the digitization of the media and the creation of the digital resource, are entered by annotators when the SLAAP record is created.

---

3 Again, I do not discuss IRB and access issues here. Readers are referred to the *SLAAP User Guide* (2008) for information about how SLAAP manages permissions and implements access control.
Figure 2.2.1. Assorted screenshots from SLAAP, discussed in the text (from Kendall 2007a).
Finally, this view provides overview information about the other accumulated materials within SLAAP for the interview and its media files, including links to variable tabulation sheets and transcripts as well as a graphical timeline of each media file, showing where transcripts exist for the audio and where users have made time-stamped notes. Beyond the information for each interview, SLAAP also provides overview information for the speakers and projects in the archive. For example, users can view each speaker’s record, as in Figure 2.2.2, and have immediate access to the interviews in which that speaker participated, the tabulation sheets available for the speaker, and so on.

![Image: SLAAP screenshot of speaker record for PEO, from Princeville, NC](image)

**Figure 2.2.2.** SLAAP screenshot of speaker record for PEO, from Princeville, NC

In addition to having immediate access to the data in the archive through the web-based interfaces, SLAAP also provides simple tools that allow for the better organization
and maintenance of additional materials. Users can upload and store files of various types in the system, such as legacy transcripts in MS Word format or photographs of locales. These can be associated with specific interviews or media files, or alternatively with individual speakers, research projects, or field sites. At the speaker-level, the system also allows for the uploading and association of vowel-plots, as illustrated in Figure 2.2.2 above (cf. Thomas 2001), which are then displayed inline as a part of the metadata available for the speaker.

At the project-level, the system allows for the association and storage of publications and manuscripts that have come out of or relate to each project. This is illustrated in Figure

![SLAAP screenshot of project overview page, showing information for the NCLLP's Hyde County research site](image)
2.2.3, a screenshot of the project overview page showing the record for the NCLLP Hyde County project (cf. Wolfram and Thomas 2002). These features enhance the long-term storage and management of project-, speaker-, and interview-related notes and data, ensuring that there is a central, secure location for the storage of the diverse materials that are acquired and created in the research process.

Analytic features in SLAAP include: (5, in Figure 2.2.1 above) tools that aid in the extraction and tabulation of linguistic variables (a close-up is provided in Figure 2.3.3, in the next section), phonetic analysis features, and (6) sophisticated transcript options. Transcript data are linked to the audio files and transcripts can be viewed in a number of formats at the same time as one listens to the associated audio (see also Figure 2.3.1, in the next section). A version of Praat, the open-source phonetics software (Boersma and Weenink 2007), is integrated into the SLAAP software to allow for the instantaneous retrieval of phonetic data (such as pitch and intensity readings) as well as the generation of spectrograms in-line with the transcript text (see Figure 2.3.2, in the next section). Finally, SLAAP offers corpus-like tools (such as (7)) that allow for large-scale linguistic analysis across interviews, speakers, and research projects. Some of these “corpus-like” tools are used and discussed in the Part 2 of this work.

A full overview of SLAAP’s features, including many screenshots, is available in the form of a user manual (SLAAP User Guide 2008). The most up-to-date version is available as a PDF at http://ncslaap.lib.ncsu.edu/userguide/.
2.3. Some benefits of explicit sociolinguistic data management work

By digitizing the entire NCLLP collection and incorporating the recordings into a centralized repository, we have in a sense put into dialogue our entire collection. Our descriptive metadata – i.e., the information stored about each interview, speaker, and research project – along with our transcripts and researcher notes are all searchable both within and across projects. Older materials and metadata are just as easily retrieved as new materials. This explicit management work creates a level of organization that is more complete and useful than otherwise. It makes for better analyses by giving us easier access to our data. It makes it easier to collaborate on research projects and share data and findings, and to do this with greater geographical distance between investigators (as Mallinson and Kendall forthcoming implicitly illustrates). Additionally, it can also create opportunities to evaluate new research questions. The analyses of Part 2 are made possible on the one hand by SLAAP’s software and data model, but also on the other hand by the fact that the recordings from disparate studies are brought together and easily compared.

In Chapter 1 I briefly problematized the analytical practices of transcription and variable tabulation. I return to these topics now to address how SLAAP has attempted to strengthen these foundational sociolinguistic methods.

2.3.1. Re-examining transcription
SLAAP seeks to apply standard data management and presentation methodologies to the treatment and representation of transcript information. One major premise therein is the separation of content and format. Separating the transcription from its formatting provides a huge amount of flexibility in terms of the presentation of the information. Following Edwards’ (2001) terminology, the same transcript can be viewed in a *vertical* format (as in (1) in Figure 2.3.1) or in *column-based* format (as in (2) in Figure 2.3.1; cf. Ochs 1979), or even what is referred to in SLAAP as a *paragraph* format (as in (3) in Figure 2.3.1).

Alternatively, that same transcript can be displayed in a purely visual format (as shown in (4) in Figure 2.3.1). In this view, called a *graphicalization*, speakers’ utterances are situated within the complete interaction in a way that gives analysts a simple visual...
overview of the unfolding of the speech event. Each speaker’s talk is displayed on its own tier. Shading indicates speech rate, with darker shading indicating faster speech, and pauses and speaker overlap are accurately depicted. Analysts can “mouse-over” utterances to see the transcript text and can click on a passage to move to deeper analytic views of the transcript (as discussed momentarily, and shown below in Figure 2.3.2).

Transcript data in SLAAP are stored in database tables. Each transcript is a table in the database, and each line is an entry in the database table representing an utterance by a speaker. Transcripts for SLAAP are built using Praat to obtain highly accurate start-and end-times for each utterance. Unlike the textual accuracy that many transcript theorists aim for, SLAAP transcripts target temporal accuracy with the belief that everything else can be (re-)constructed from the audio file, either automatically by software, or manually by examining the audio for the given time range.

Chapter 3 discusses transcription in further depth. Here, though, note that in a data-based transcript model, the only data required for a complete transcription unit are: (a) a reference to which speaker in the interaction is speaking, (b) the utterance’s start time, (c) an orthographic representation of the utterance, and (d) the utterance’s end time (Kendall 2005, 2006-2007, 2007a). This very simple data model is actually quite powerful. Speech data management software, like SLAAP, can then create links between the transcript data and the audio file from which the transcript is based, and phonetic

---
4 SLAAP’s method for calculating speech rate is discussed in §6.4.
5 See also Barbiers, Cornips, and Kunst (2007) on using Praat for time-aligned transcription. MacWhinney (2007) discusses some other tools that can be used for this purpose, such as Transcriber, and CLAN. Kendall (2005, 2006-2007), MacWhinney (2007), and others (e.g. Edwards 2001) provide general discussions on the benefits of time-aligned transcripts.
software (such as Praat in the case of SLAAP) can be integrated with the transcript to allow for real-time phonetic analysis. In other words, with the start- and end-times for each utterance captured in the database and a linkage maintained with the audio, much of the other information that is often tagged or coded (e.g. latching, overlap, pause length, etc.) is unnecessary and can be reconstructed from the audio itself (as demonstrated in the graphicalized view shown in Figure 2.3.1).

At the same time, an approximation of standard orthography (cf. Chafe 1993: 34; Tagliamonte 2007: 211-5) is sufficient for the transcript text because pronunciation features (e.g. vowel qualities, r-vocalization, etc.) can be listened for or examined instantly via a spectrogram. This simple orthography makes the transcripts easier to read than more complex systems, especially for new readers and non-experts. The use of standard orthography also allows for easier searching through SLAAP’s search features.

Figure 2.3.2 shows a screenshot from the SLAAP software demonstrating an in-depth view of one transcript line. This example shows a pitch plot as well as a spectrogram, though other data-views are available. Note also that the audio for the line can be listened to through an embedded audio player and that numerical data (in Figure 2.3.2 acoustic measurements of pitch) can be obtained at the click of the mouse.

Additionally, multiple transcript lines can be displayed in this detailed format on the same page, allowing for the easy comparison between utterances or individual word-tokens.

---

6 Also see Preston (1982, 2000) on the importance of the choice of orthography in transcription.
Another major benefit of the SLAAP approach to the treatment of natural speech data is that quantitative and qualitative analyses can be better integrated with one another. With linkages maintained between the quantified data and the speech events from which the data are extracted, analysts can better situate their quantitative data and analyses in terms of the larger discourse. Likewise, discourse-level work, typically focused more on qualitative questions, can more easily integrate quantitative measures. This section seeks to illustrate some of these benefits by re-examining variable tabulation through its implementation in the SLAAP software.

SLAAP’s variable tabulation tool helps to counter some of the problems highlighted earlier in this essay by making tabulation practices more transparent and
individual tabulation data more accessible for easy review. Following the focus on temporal accuracy behind transcription implementation (as discussed above in §2.3.1), tabulations in SLAAP are time-stamped entries linked to the source audio. As with “normal” variable coding practice, each tabulation consists of a set of enumerated fields. To “extract” an occurrence of a variable, an analyst simply clicks a button to retrieve the timestamp from the audio player, enters a short string of text describing the context of the occurrence, and then selects the appropriate attribute from a pop-up menu for each parameter. For illustration, Figure 2.3.3 shows a part of the variable tabulation screen in SLAAP for the variable syllable-coda consonant cluster reduction (CCR; cf. Wolfram, Childs, and Torbert 2000, or Guy 1980 on the related variable of t/d deletion).

Figure 2.3.3. SLAAP screenshot of a tabulation form (for CCR) and audio player
Through the linkage with the source recording, analysts are able to review their own tabulations by returning to the appropriate moment in the audio at the click of the mouse and colleagues can easily share and review each other’s work. Furthermore, coding analysts are prompted to mark their level of confidence for each tab, which provides a helpful mechanism for the review of putative or less confident tabulations. Since tabulations are time-aligned to the source recording, the system maintains a connection between these quantified data and the greater context of the interaction. Since quantitative variable data are time-stamped in the same way as the transcripts, variable data can be overlaid upon the transcripts – either in the text-based or graphical views. In other words, tabulation data remain situated in terms of the larger discourse, enabling more holistic analyses of the data (such as examining the role of topic, stance, or interlocutor effects on variable productions over the course of the speech event). In Chapter 10, I make use of the time-stamped variable tabulation model to examine the relationship between variable realizations and the temporal sequencing patterns of their matrix talk.

In addition to the benefits of coding transparency, improved accuracy, and better-situated quantitative data, this method also provides simple logistical benefits. Through the web-based interface, analysts can tabulate their data from any Internet-connected computer and can leave their work and return to it without losing their place in the audio. As shown in Figure 2.3.4, SLAAP also allows users to view tabulation summary results directly from the website as well as to download tab-delimited versions of the tabulation.
sheet suitable for opening in Microsoft Excel or other spreadsheet applications. In sum, the procedural enhancements provided by the SLAAP implementation of variable tabulation enable general methodological and theoretical advances to this foundational component of quantitative sociolinguistics.

2.4. **SLAAP’s data model**

In terms of actual technologies, SLAAP uses a MySQL database server (http://www.mysql.com/) running on Apple Macintosh X Server hardware. SLAAP is
served by the Apache webserver (http://www.apache.org/) and most of the software is written in PHP (http://www.php.net/). While much corpus work has been undertaken using XML or other markup schemas (cf. Bird and Simons 2003; Simons, Fitzsimons, Langendoen, Lewis, Farrar, Lanham, Basham, and Gonzalez 2004), the decision to use a relational database system (MySQL) was made based on the hypothesis that linguistic data could be beneficially treated in simplistic data terms (Kendall 2006-2007; Kendall and French 2006) and that processing speeds and operational simplicity would outweigh the flexibility of an extensible (XML-based) mark up schema. Davies (2005) also makes a compelling case for the use of relational databases for linguistic corpora.

Operationally, SLAAP seeks to provide its users better tools and better data with which to undertake their studies, whether traditional sociolinguistic pursuits or investigations of new avenues for research. As such, the project has been presented here as an example of ways in which sociolinguists can move their data collections from “tapes in a cabinet” to interactive and powerful tools for linguistic analysis. This sort of interactive archive increases the utility of speech data. The steady accumulation of data and metadata in a corpus such as this – researcher’s notes, transcripts, variable tabulations, and so forth – enhances the collection overall. Instead of data becoming less usable over the course of years (as the original analysts move on, notes are misplaced, the audio tapes deteriorate, etc.), the speech data stored in a system like SLAAP become richer and more usable.

---

7 The server is housed by the NCSU Libraries. Other than the operating system itself (Macintosh), all of the server software are freely available open-source technologies.
Figure 2.4.1. Re-conceived model for abstraction in sociolinguistic data (from Kendall 2008a)
Through a speech data management system like SLAAP and its ability to link between levels of metadata, we can re-conceive the schematic first depicted in Chapter 1 as Figure 1.4.1. Figure 2.4.1 (from Kendall 2008a) presents this improved schematic, as is conceptualized within SLAAP, alongside the original illustration for comparison. In the re-conceived version, metadata types are closer together, and are connected in all cases with bi-directional arrows that depict the stronger connection between the levels of representation. Thus, metadata are interconnected, each making the others richer, more accurate, and more fully understood. There are also fewer levels of abstraction, because of the iterative connection between the base-data of the recording and each level of metadata.

Of course, a system like SLAAP cannot remove all abstraction from the representation of linguistic information (hence the not-completely-solid lines); it is not possible to do so entirely. However, as SLAAP illustrates, levels of metadata can be moved closer to their source recording by maintaining strong linkages between the two. A system such as this also provides a level of organization and some basic shared methodologies for all of its users. Thus, it makes more explicit the data management and analytic practices that operate on the data within the archive and, through features like SLAAP’s variable tabulation implementation, it can contribute towards making sociolinguistic data more robust and analyses more reliable.

2.5. In closing: Strengthening our relationship with our data
In closing, the goal of this chapter – and, ultimately, of SLAAP – is to highlight some of the ways in which sociolinguists can improve the organization, storage, accessibility, and preservation of their data collections, both for their current work and for future work with those collections. SLAAP instantiates one approach to improved sociolinguistic data management, and, ultimately, demonstrates that regardless of the specific model adopted, it is time for sociolinguists to reconsider and make more rigorous their data management practices.
3. Implementing a databased transcription model

3.1. Transcription as theory and practice

Although transcription has already been mentioned in the previous chapters, there are a number of issues both within the theory of transcription and within the particular model for transcription implemented in SLAAP that warrant a more extended discussion. Further, SLAAP’s transcription model also bears on the data and analyses of Part 2 and a deeper discussion is hoped to benefit readers principally interested in those analyses.

Thirty years ago Elinor Ochs (1979) published the keen observation that the way we put speech down on paper as written words is far from theory-neutral.¹ From a focus on the pragmatics of child language, Ochs argued that, until that point (and excepting some of the work within conversation analysis; Sacks, Schegloff, and Jefferson 1974), transcripts were often used as a major component of linguists’ data, but that transcription is a selective process, and, as a process, had not been explicitly or adequately treated by empirical linguists (1979: 44). Ochs explained,

a more useful transcript is a more selective one… Selectivity, then, is to be encouraged. But selectivity should not be random and implicit. Rather, the transcriber should be conscious of the filtering process. The basis for the selective transcription should be clear (Ochs 1979: 44).

¹ Readers will note that a part of the title for this section, “Transcription as theory” is taken from the title of Och’s important (1979) paper.
She went on to explain that by ignoring the theoretical and “selective” aspect of transcription, in fact, “researchers rarely produce a transcript that does reflect their research goals and the state of the field” (Ochs 1979: 45, emphasis added).

In the decades that have followed, many others have echoed this important connection between transcription and theory (e.g., Jefferson 1983; Mishler 1991; Bucholtz 2000; Edwards 2001). It is clear that we take her words seriously. In fact, I do not believe that many scholars would still consider transcripts to be their data in as direct a sense as Ochs claimed for the time, partly thanks to her early work along these lines.

So, for example, Johnstone’s (2002) textbook, Discourse Analysis, states that discourse analysts “study records of discourse [and] these records are often in the form of transcripts of audio- or videotapes” (Johnstone 2002: 19, emphasis in original), but then immediately spends the next few pages explaining that

any analytic move that involves drawing boundaries, pulling out chunks from the flow of experience and treating them as wholes, is somewhat artificial (20).

In short, Johnstone nuances the notion of transcript-as-data to stress at length the theoretical nature of transcription and the fact that there is no essential relationship of transcript = data.

Even scholars who are not invested specifically in the literature and theory on transcription and/or discourse analysis seem to readily acknowledge these points (cf., e.g., Tagliamonte 2006, 2007 on variationist sociolinguistics; Gibbons 2003 on forensic linguistics). Despite this fact, and the increasing focus on improving transcription conventions (cf. Du Bois, Schuetze-Coburn, Cumming, and Paolino 1993; Du Bois 2006).
and on incorporating time-alignment into transcription practice (cf. Edwards 2001; Kendall 2005, 2006-2007; MacWhinney 2007), few far-reaching methodological improvements have been widely adopted when it comes to standard transcribing protocols. That is, much transcription work is still done in word processors (such as MS Word) and built around the needs of a particular, individual research project. Often times the central concern when it comes to the transcription stage of a research project is to complete the transcription work as quickly and cheaply as possible to satisfy the needs of the current project, without fully considering the benefits (and drawbacks) of various options for transcription.

This is clearly true in areas outside of linguistics, such as in the legal profession, where transcription practices have proceeded in a business-as-usual way, despite the recognition by and arguments of language and law experts that current transcription practices obscure much about actual spoken language. It has also been acknowledged that they can create bias in the legal system by representing some speakers as hyper-articulate and others as less educated (Walker 1990; Gibbons 2003). Transcripts can also be misleading in their portrayal of the passage of time. In Kendall (2007b), I examined the legal transcript of the deposition of an expert witness along with the video recording of the deposition in order to assess how well, and more generally just how, the occurrences of the deposition were captured by the professional legal transcriber. I found that the transcript was potentially misleading, especially in terms of its accounting for the

\[\text{2 The same has been said about transcripts more broadly, that certain portrayals of dialects, such as forms of eye-dialect, misrepresent speakers more than they represent actual aspects of dialect (Preston 1982, 2000).}\]
passage of time and its failure to accurately indicate not only pauses but extended passages of time during which the deposition participants conducted non-verbal action. These sorts of inaccuracies can have far-reaching implications on the interpretation of the transcript and can lead to failures in the justice system (Butters and Kendall submitted).

In short, the legal profession’s focus on the *verbatim* transcript (cf. Gibbons 2003) fails to provide the users of the transcript with crucial information about non-verbal activity in the deposition. While there was a video record of the examined deposition (a common practice for depositions) for the case I examined in Kendall (2007b), the lack of time-alignment within the transcript even made the transcript minimally useful as a guide to the recording.

The above digression into language and law is mentioned here primarily to lend support to the argument of §2.3.1 that transcription should be re-oriented towards a focus on temporal accuracy and that there are far reaching benefits beyond sociolinguistic analysis when transcripts are oriented thusly. At the same time, this is not entirely to argue *against* textual accuracy as an important goal for transcription, but instead it is to argue that textual accuracy can be de-emphasized when greater temporal accuracy is achieved.

For all but the clearest recordings, transcribers will no doubt disagree about certain utterances – e.g., how a speaker pronounced a particular word, or what word was spoken. However, with the proper instrumentation, transcribers can obtain rather high agreement in determining the boundaries of speech. That is, the time-alignment task
seems to be a more reliable task than the actual transcription task.\textsuperscript{3} With the right kind of tool (e.g., SLAAP), it seems advantageous and maximally flexible to push back the concern with orthographic accuracy until as late in the analytic process as possible. In other words, we should embrace the aspects of transcription that we can conduct with accuracy and instrumental rigor, instead of focusing our energies on the exact specifics of orthographic decisions.\textsuperscript{4}

The argument for an emphasis on temporal accuracy – that is, time-stamped text that is linked to the audio – raises an important question for the transcription process: at what level of granularity should (or can) the transcript be time-aligned? In other words, what is the best temporal unit for a time-aligned transcript? We rephrase these questions for the next section.

### 3.2. What constitutes a transcript line?

I here use the term *transcript line* to denote a unit in transcription. In SLAAP’s time-aligned transcription system, lines are temporal units, but in other transcription systems and in the sorts of transcript excerpts often included in publications for illustration, transcript lines may be units of a different sort. Regardless of its format (cf.

\textsuperscript{3} I have not quantitatively assessed inter-analyst comparability for Praat-based transcript time stamping, and this surely should be done. However, a number of transcribers (from professional researchers to undergraduate students) have contributed transcripts to the SLAAP archive and after even just a little training the resultant transcripts appear quite comparable. This is discussed in slightly further depth in §4.4.

\textsuperscript{4} SLAAP allows users (with proper access privileges) to edit the transcription orthography directly from the transcript. This allows users to amend transcripts in situ, and further reduces the need for textual accuracy in the initial phases of transcription.
Ochs 1979), there is a need to consider a transcript as composed of parts and each of
these parts – whether representing a speaker’s turn, a clause or other syntactic unit, an
intonation phrase, or a stretch of uninterrupted phonation – can be considered a *line* for
the purposes of transcription. So, what does it mean when a transcriber ends a line and
starts a new one? These lines are often numbered and used as a way to index the
transcript. This is clearly useful for referencing a location in a transcript for colleagues or
in publication, but it is often the case that in analysis and discussion individual transcript
lines are treated as units of data. But, what exactly are they units of?

In §2.3.1, I explained that SLAAP implements a databased transcription model,
wherein each transcript is stored as a database table in a relational database. Each
transcript line is stored as a row in its transcript’s table. At that point, I explained that a
transcript line in SLAAP is a phonetic utterance – speech surrounded by silence on the
part of its speaker. However, I did not explain why that definition of a transcript line was
used when there are numerous ways that we can segment discourse into units. We turn to
this question here.

That this question – the problem of segmentation – has not been satisfactorily
answered is illustrated by the fact that the Linguistic Data Consortium (LDC), the largest
creator and distributor of text and audio corpora, appears to use a variety of segmentation
schemes despite their expertise in annotation and linguistic data creation. Yuan,
Liberman, and Cieri (2006), for example, in discussing the segmentation of a number of
LDC corpora, describe the corpora as segmented into “pause groups”, but point out that
these segments are often equivalent to speaker turns, and may be regarded
as a reasonable proxy for turns, though longer turns may be divided into
several segments… Unfortunately, the procedures for creating these segments are not consistent across (and sometimes even within) the cited corpora (2006: 1).

In other words, there is *often* an equivalency in transcription units, but why only *often*?

In defense of the LDC, segmentation in corpora is most often determined based on the particular purpose of each corpus, and just because corpora are published by the same organization does not imply that the corpora are published for the same purposes. At the same time, however, the lack of consistency *within* corpora indicates that at least some of their segmentation protocols are, perhaps, too subjective. The overall fact that there is variation in the segmentation of corpora is not really so surprising, however, because it relates, of course, to the larger question of defining and determining a *unit* of speech, and it is not clear that we can answer this question sufficiently, in absolute terms and for all purposes. In Levelt’s (1989) words, “there is no *single* unit of talk” (23, emphasis added). Common practice, and my own practice, is to call this (at least imagined) talk-unit an *utterance*, but, then, what is an utterance?

### 3.2.1. What is an utterance?

On the one hand,

an utterance is more than a sentence, i.e. an utterance is more complex than a sentence having both linguistic and non-linguistic properties and functions. The abstract notion of the sentence is insufficient to account for utterance. It is therefore not surprising that the sentence as a syntactic unit has been shown to be inadequate for the study of discourse, whether by discourse is meant language in context, language use, or language beyond the sentence level (Figueroa 1994: 163).
At the same time, an utterance, as a unit of talk, is often less than a sentence. In a number of publications, Wallace Chafe has explored this question and argued for a notion of unit called variously spurt (Chafe 1980b, 1985), idea unit (Chafe 1980a), and intonation unit (Chafe 1994). While his terminology has changed over the course of his research, the basic premise and finding – that these spurts have a cognitive basis and are rooted in the consciousness of the speaker (cf. Chafe 1994) – has only strengthened. Chafe finds that “the majority of substantive intonation units have the form of single clauses, though many others are parts of clauses” (1994: 69). Talking about the same data in an earlier work, he explains:

In the data available to me at the moment these spurts are slightly less than 2 seconds in mean duration, and contain about 5 words. They tend to be single clauses syntactically, but under certain conditions may be more or less than a clause (Chafe 1980b: 171).

This possible conceptualization of a talk-unit is just one of many, however. Levelt (1989: 23), for example, readily lists 18 different terms that have been used for speech units and points out that his list could easily be tripled. To consider some other units of discourse that have received traction in the literature it seems appropriate to mention the system(atic) proposed by Sacks, Schegloff, and Jefferson in their foundational (1974) paper. This work veritably created the field of conversation analysis by moving an analytic lens to the very specific workings of communicative interaction. In many ways their work, and the body of work within conversation analysis that followed, has great bearing on the definition of a unit of discourse. Sacks et al. attempt to
understand the ways in which interlocutors themselves determine turn boundaries in ongoing speech.

Sacks et al. (1974) point out that there are various “unit-types” which can be used to construct speaker turns at talk, including (in English) “sentential, clausal, phrasal, and lexical constructions” (1974: 702). They call these units turn-constructional units (TCUs). Importantly, they note that instances of a unit-type are identifiable because they project their type, and “what, roughly, it will take for an instance of that unit-type to be completed” (702). Listeners, in other words, are for the most part able to predict the ends of their interlocutors’ turns because of the projectability of turns. Sacks et al. also point out that these unit-types are all to a certain degree syntactically-based – that “syntax matters to turn-taking, albeit a syntax conceived in terms of its relevance to turn-taking” (1974: 721). However the relevance of this syntax to turn-taking appears to be possibilized via intonation (722) and in later considerations (e.g. Schegloff 1996; Selting 2000) prosody appears to play an increased role in the determination and definition of the TCU.

More importantly, as Selting (2000: 478) clarifies,

The TCU is thus a “unit” in conversation which is defined with respect to turn-taking: a potentially complete turn. The TCU is not defined as a linguistic unit.

While the construal of the TCU as a non-linguistic unit made sense for the needs of conversation analysts, for the purposes of the current discussion – an assessment of the most meaningful practical definition of an utterance – I would argue that we are clearly interested in a unit that is linguistically defined or linguistically definable.
In fact, both this increased emphasis on intonation and the fact that the turn-constructional unit does not seem adequate for a definition of a talk-unit on linguistic grounds returns us to Chafe’s work and the intonation unit. While a full review is outside the scope of this dissertation, Chafe and his colleagues (e.g., Du Bois 2006) make a compelling case for the intonation unit as a central unit of speech and useful for transcription. However, recalling Levelt’s short quotation above (1989: 23), the intonation unit is not the central unit of speech and we can actually consider the work on intonation units as best situated as just one case in the search for the minimal discourse unit in a larger literature and array of research (cf. Degand and Simon 2005).  

Schuetze-Coburn, Shapley, and Weber (1991) provide an example of this in their in-depth comparison of auditorily defined intonation units (of the Chafe sort) with acoustically determined declination units and show that while the two are related there is not a 1-to-1 relationship. Declination units – based on an acoustically determined pitch-reset and declination cycle – were longer than intonation units in their data, with an average of 1.8 intonation units per declination unit. In fact, their work indicates that the specific acoustic features of DUs [declination units] (F₀ reset and pause) had perceptible auditory correlates which the auditory analysts were consistently identifying when segmenting the corpus into IUs [intonation units] (Schuetze-Coburn et al. 1991: 231).

However, the opposite was not found to be true; the intonation units were not found to correlate fully with any of the acoustic features examined. While Schuetze-Coburn et al. draw from their analysis a “justification to the applicability of auditory data to acoustic analysis” (231), they also show that the use of acoustically determined declination units,  

---

5 I thank David Herman for talking with me about this section.
“while lacking the detail afforded by the IU analysis, adds a measure of cohesiveness not available from IUs” (231). In short, they provide evidence that to me is in favor of an acoustic definition of an utterance.

For the present discussion, one area that is also potentially problematic with respect to the intonation unit as the primary unit of talk is that intonation units are often incomplete in actual talk. Chafe himself notes that there can be unsuccessful intonation units, which he terms *fragmentary* intonation units (1994). So, while I am convinced of the importance of the intonation unit at many theoretical levels and, I believe, there is some acoustic evidence for the coherent nature of an intonation phrase (Schuetze-Coburn et al. 1991: 225), I also wonder whether the “tune” of speech is really anything more than a grind of our laryngeal engine (cf. Pierrehumbert 1980). We are able to stop mid-cycle, to pause, to restart or reframe our speech. And this fact, at least methodologically, tempers the universality of the intonation unit (or the declination unit) as *the* minimal discourse unit.

Chafe notes that intonation units (*idea units* in his 1980a terminology) are generally delimited by a combination of “intonational, hesitational, and syntactic” cues. However, “all three are not always present, nor does the presence of any one of them necessarily signal the boundary of an idea unit” (Chafe 1980a: 14), and he argues that “clause-final rising or falling pitch is the single most consistent signal” (14). That is, intonation is the most important factor.\(^6\) Pauses, he says,

frequently occur within idea units as well as between them. In other words, although the intonational and syntactic criteria associated with idea

\(^6\) Hence, his later renaming to *intonation units*.  

56
units are more often than not accompanied by at least a brief pause, the mere presence of a pause is not a sufficient indicator of an idea unit boundary (15).

While Chafe argues that “in most cases the identification of idea units is not a difficult task, and problematic cases are relatively few in number” (15), the fluidity of the three criteria – intonation, pause, and syntax – and the necessary subjectivity required to delimit utterances in these terms led me, for my work on SLAAP, to adopt the hypothesis that a useful working definition for a unit of speech might best equate with stretches of phonation, separated by measurable silence on the part of a speaker. That is, I hypothesized that defining an utterance based on a single required, acoustically measurable factor would be methodologically useful, as it allows a simplistic definition and instrumental precision in the delimitation of speech.

This conception of an “utterance” has also been proposed elsewhere (see, for example, Harris 1960, mentioned in Figueroa 1994). While Goldman-Eisler (e.g. 1968) was not concerned with transcription theory, her interest in pauses and their role in speech production and perception also led her to a “phonic” conception of utterance:

If we measure vocal continuity by the number of words uttered between two pauses and call ‘phrase’ the sequence uttered without break… (Goldman-Eisler 1968: 16).

This parallels my own thinking and I wonder whether our shared interest in pause – and the realization that pause comprises such a large portion of actual talk – led us to this shared conception. As indicated in the quotations from Chafe above, there does appear to

---

7 Surprisingly, Goldman-Eisler’s talk about “utterance” was fairly imprecise. For example, she defines “utterance” at one point thusly: “utterance here refers to a period of speech sandwiched between the last word of the interlocutor’s preceding utterance and the first word of his following one” (1968: 18). Overall, it is not clear whether she was using “phrase” and “utterance” as interchangeable.
be some evidence that phonetic units and intonation units overlap in many cases. Degand and Simon (2005; citing Mertens 1993) also point out that “major intonation boundaries … are usually signaled by the presence of silent pause.” Schuetze-Coburn et al. found that 96% of the measured pauses longer than 300 ms in their data “preceded either acoustic or auditory unit boundaries or both” (1991: 222) but also that both kinds of units were possible without preceding pauses.

Stepping back, Levelt reminds us that “much ink has been spilled on the question of what units of processing are involved in speech production” (Levelt 1989: 23). Further, he points out “the empirical evidence marshaled for one unit rather than another has been very diverse, including pause patterns, intonational structure, speech errors, and speech-accompanying gestures” (Levelt 1989: 23). Figueroa (1994) further points out at length the difficulties of even talking about “utterance” (cf. 1994: 173, footnote 1). For the sake of the present project, we end this discussion with the acknowledgement that there are many possible conceptions and definitions for units of speech and that, on account of the diverse options, we might be benefited by taking as simple an approach as possible for implementing a unit in SLAAP.

3.2.2. Motivating the distinction

The adoption in SLAAP of the phonetic utterance as a unit of speech is importantly motivated by methodological utility and not theoretical importance. In other words, this is not to argue that the phonetic unit is a cognitive reality, or otherwise is
bound up with language production. In fact, the adoption of unit qua phonetic utterance is, more than anything else, a convenient mechanism; it is methodologically useful. While Chafe argues that intonation unit delimitation is achieved with high levels of inter-analyst/inter-transcriber agreement (1980a: 15), the interpretation of intonation remains a relatively impressionistic task. Even if acoustic measures are determined that can lend instrumental precision to the task (cf. Schuetze-Coburn et al. 1991), pitch tracking software – like that available in Praat (Boersma and Weenink 2007) – can have trouble tracking pitch in noisier audio files (often times the sort with which we, sociolinguists, work).

The determination and delimitation of pauses, on the other hand, can be undertaken with higher levels of instrumentally derived accuracy. SLAAP’s archive is designed to grow organically, with minimal editorial oversight. Users should be able to add to the archive – in particular, to add transcripts to the archive – easily and to be able to use their transcripts immediately. As such, it is necessary to have a simple set of transcript conventions, and to ensure that for the “pieces” that really matter (in SLAAP the timestamps) there is a clear, simple, and accurate way for all users to generate relatively comparable transcripts.

Many of the features in SLAAP have been designed around maximizing the utility of the transcript-media connection through what is possible with current technologies. I have not focused on trying to develop (or waiting for) natural language processing applications beyond what is possible now. So, for example, SLAAP’s software is based on entirely human transcribed audio – there has been no attempt to incorporate or
develop, for example, speech recognition functionality in the software. That said, the adoption of the phonetic unit as a transcript line does support the potential automation of speech segmentation. Of all the natural language processing endeavors that are being investigated, the automatic segmentation of speech around the silence-speech-silence phonetic unit is one of the most obtainable goals. In fact, for recordings of high-enough quality (probably better quality than many of the field recordings in SLAAP), pause detection could be automated, even through relatively simple scripts written in Praat.  

In any case, this section has sought to provide some rationale for the adoption of the phonetic utterance as the primary unit of transcription within SLAAP. The decision to focus on the phonetic utterance for SLAAP transcripts has enabled the analysis of speech rate and pause (in Part 2) in more rigorous and systematic ways than could have been possible otherwise. As discussed shortly, in §3.4, it also makes possible new sorts of visualization and presentation strategies that actualize some of the improvements to transcription discussed by scholars such as Ochs (1979) and Edwards (e.g., 2001). But, first, in the next section I provide further detail about the transcription process and discuss the major conventions used in SLAAP transcripts.

### 3.3. Transcribing for SLAAP

---

8 Provided sufficient quality recordings, the hardest problem for automatic speech segmentation is the determination of which speaker to attribute a segment of talk.
Although Praat (Boersma and Weenink 2007) was not designed to be a transcription tool, it is, I believe, the best tool for the transcription job, especially when accurate time-alignment is intended. Praat is free, open-source, and works on all the major computer operating systems.\(^9\) It also allows for multiple levels of annotation or transcript, beyond those used here (cf. Vaughn 2008). Finally, and most importantly, Praat allows for arbitrarily fine-grained time-alignment for any number of speakers.\(^{10}\)

I originally intended to give here a technical “how-to” guide for the transcription process for SLAAP using Praat. Instead, that information is now available online in the SLAAP User Guide (2008). I here briefly outline the process so that readers understand the methodology used in generating SLAAP’s transcripts, since they are central in enabling SLAAP’s features described in Chapter 2, and form a major component of the data used in the analyses of Part 2.

3.3.1. *The TextGrid and TextTiers*

Transcription in Praat takes place using the TextGrid annotation object type. A transcript is made in a TextGrid and the speech for each speaker in the transcribed

\(^9\) Of course, other transcription software programs available (such as Transcriber; Barras, Geffrois, Wu, and Liberman 2001, available at http://sourceforge.net/projects/trans/) are also free, open-source, and cross-platform. However, they do not provide the fine-grained time-alignment possibilities enabled by Praat. As an aside, I have made available a public tool – at http://ncslaap.lib.ncsu.edu/tools/trans_to_praat.php – that converts Transcriber transcript files to Praat readable files.

\(^{10}\) In fact, I recommend Praat for transcription for everyone, beyond SLAAP users. A public tool is available on the SLAAP website – http://ncslaap.lib.ncsu.edu/tools/praat_to_text.php – that converts from Praat format to readable text to aid non-SLAAP users.
interaction is contained in an interval tier of the TextGrid. Figure 3.3.1 shows the Praat editor window for a transcript with two speakers, along with the audio data. Typically, the tiers are ordered by interest and amount of talk. So, the top tiers will contain the main interviewee(s), below those will be the interviewer(s), and at the bottom less talkative or important speakers as well as interlopers.

![Figure 3.3.1. Screenshot of Praat editor window (media: ohdhud_m)](image)

Importantly, each tier of the TextGrid provides a complete accounting for that speaker’s speech over the course of an interview. Empty intervals indicate when

---

11 Again, this is not intended as an introduction to the Praat software or a full tutorial. See the Praat website – http://www.praat.org/ – for documentation and tutorials on using Praat and the SLAAP User Guide (2008) for more complete information on transcribing for SLAAP.
speakers are silent, while text in the intervals transcribes the speakers’ talk. In their current conception, SLAAP transcripts do not attempt to record or describe non-verbal action, although in general parentheses within the relevant tier are used to record notes about a speaker’s talk or action (see next section).

The most important aspect of a time-aligned transcript is, of course, the accurate delimitation of the speech timing. All pauses greater than about 60 milliseconds are delimited as pauses, by being marked off with interval boundaries and with no text within the interval. The alternation of utterance and pause is visible in Figure 3.3.1, above.

3.3.2. Orthographic conventions

One of the primary utilities of a time-aligned transcript – at least within its conceptualization within SLAAP – is to act as a proxy to the original recording, providing a means for easy searching and browsing of the recording. It is not necessarily to make a textually accurate representation of the speech (if that is even possible). Along these lines, the transcript often uses simple orthography and standard-like spelling. In general, morphosyntactic variants (e.g., was for were) are transcribed, but phonological variants (such as r-lessness, or Northern Cities vowel qualities) are not.\(^\text{12}\)

\(^{12}\) While I outline here the orthographic conventions that I use and recommend, the power of the temporally accurate transcript, and its implementation in SLAAP, is that users have wide latitude in their orthographic decisions. Some transcribers put more emphasis on textual or phonological accuracy than I do. These transcription differences do impact the searchability of the overall archive, and the ability in certain cases to compare lexical types through corpus-based methods and features (not discussed here; see the *SLAAP User Guide* 2008), but do not impact the features or methods use in Part 2’s analyses.
At the same time, the transcript text attempts to accurately account for all the “noises” of speech, such as laughter, filled pauses (like “uh” or “um”), restarts (e.g., “I- I di- didn’t mean to”), and misspoken words (e.g., “brack in the seventies”). Standard-like punctuation is often used, with the hyphen, -, used to indicate lexical and morphosyntactic restarts, as well as incomplete intonation. Silent pauses (of course) are never described or coded, as they are represented in all cases by empty intervals.

While the orthography of SLAAP’s transcripts is flexible in general, three features are crucially coded for, which have special characters and implications on the features used in Part 2’s analyses. These features – speaker overlap, unintelligible talk, and non-/meta-linguistic noises – are described in Table 3.3.1. Additionally, transcribers often include notes about their transcripts, and about interesting speech features found in the speech. The convention for this is also included in Table 3.3.1.

### Table 3.3.1. SLAAP transcription conventions, special characters

<table>
<thead>
<tr>
<th>Feature</th>
<th>Special Chars</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Overlap:</strong> Speaker overlap is noted by the use of <strong>square brackets</strong>, for all parties to the overlap. The overlap markers are, however, only placed at word boundaries. Since utterances are accurately time-aligned through the tier boundaries, these markers are aids for readers and not critical for timing determinations. Therefore, they are not (highly) accurately placed.</td>
<td>[ ]</td>
<td>RR: Habitat [for /Humanity/ PEO: [Habitat for] Humanity came prv007aa_840_1430, 94-95</td>
</tr>
<tr>
<td>2. <strong>Unintelligible/inaudible speech:</strong> Slashes are used to enclose sections of unsure transcription. Transcribers place “best guesses” within the slashes, or write /unintelligible/ for unintelligible talk or /inaudible/ for inaudible talk. For unintelligible talk of less than three syllables, transcribers also may use question marks, ?, within the slashes to indicate each syllable of unintelligible speech.</td>
<td>/</td>
<td>“Here? /unintelligible/ High school.” bee0010a_0_2756, 1588</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Well a rockweiler /really/ got” bee0010a_0_2756, 864</td>
</tr>
</tbody>
</table>
a. **Obscuration:**
Slashes are also used to obscure real names, when the use of a real name is inappropriate or not allowed, or to replace real names with pseudonyms.

```
//
   e.g.,
   /NAME/
   /Alayna/
   /Center City/
```

“How old are you, /Alayna/?”

3. **Non-linguistic/meta-linguistic noises:**
Noises like laughter, hand clapping, and throat clears are indicated by short descriptions enclosed within angle brackets. These are only used to describe actual noises, not features like voice quality (see below).

```
< >
   e.g.,
   <laugh>
   <throat clear>
```

“I was a full-time mommy. <cough>”

3.3.3. **From Praat to SLAAP**

Once a transcript is completed in Praat (for the time range of interest), it is added to the SLAAP archive and associated to its interview and media file through a webpage in the SLAAP software (see the *SLAAP User Guide 2008*). In this step, the SLAAP software parses the Praat TextGrid file and converts it to a database table storing it in the SLAAP database. Once the transcript is in the SLAAP archive, it can be viewed and interacted with in various ways, from the displays shown in Figure 2.3.1 to those discussed in the next section and in Part 2.

65
3.4. Visualizing speech data

As was discussed briefly in §2.3.1 and illustrated by the graphicalization in Figure 2.3.1, one of the beneficial outcomes of the sort of transcript model adopted by SLAAP, with accurate time-alignment, the maintenance of a connection between the transcript and its source recording, and integrated acoustic analysis features, is that the transcript data can be presented in any number of formats. These formats can extend beyond text-based presentations, and can be used to aid in understanding – in “reading” – our data in ways that are not possible through traditional text-based strategies.

3.4.1. Visualizing quantitative data

Visualization strategies are important and useful interfaces to complex data. As the opening words of William Cleveland’s (1993:1) *Visualizing Data* state, visualization is critical to data analysis. It provides a front line of attack, revealing intricate structure in data that cannot be absorbed in any other way. We discover unimagined effects, and we challenge imagined ones. Cleveland tells us further that visualization strategies provide us the means to take a “penetrating look at the structure of data” (1993:12), and that this “penetrating look” can be an effective analytic maneuver. Cleveland’s work is in the realm of statistical analysis, and sociolinguists without doubt make heavy use of visualization in their statistical and quantitative analyses and in their presentations of quantitative data. For example, Labov’s “crossover effect” is best illustrated by his near-famous figure from his
study of English in New York City (1966), reproduced here as Figure 3.4.1. This figure beautifully sums up the complex finding from his study that speakers in the socioeconomic status just below the top of the scale “crossover” the highest status speakers, and speak more standardly – perhaps hyper-standardly – as their attention to speech increases (Labov 1966, 1972).

![Figure 3.4.1. Crossover effect for postvocalic r in New York City (Labov 1972: 114)](image)

3.4.2. Visualizing phonetic data

Of course visualization techniques in linguistics did not begin with presentations of quantitative data. The sound spectrogram, since its increasing availability following World War II (Potter, Kopp, and Green 1947), has been a central tool within phonetic and
sociophonetic linguistic approaches. It seems fair to say that the invention of the spectrograph (the pre-computer device that could generate spectrograms) revolutionized phonetic analysis and enabled a new level of instrumental analysis of speech data. The spectrogram is so useful for analysis because it displays three dimensions of information in a highly readable two-dimensional display (time on the x-axis, frequency on the y-axis, and amplitude as the darkness of shading at a given x, y coordinate). A spectrogram, along with overlaid transcript text and pitch data, generated in Praat is provided in Figure 3.4.2 for illustration.

Figure 3.4.2. Spectrogram with text and pitch displayed

Beyond spectrograms, North American sociophonetics (cf. Thomas 2001) has popularized the F1~F2 vowel formant plot as a valuable tool for portraying – and understanding – acoustic data as well.
3.4.3. Visualizing discourse data

In sum, visualization techniques are central tools of sociolinguistic analysis, at least at the two levels discussed above – the micro-level acoustic data of the speech signal and the macro-level summative data of patterns of overall feature use. Some scholars have also experimented in displaying discourse data. For example, Podesva (2008) has developed a sophisticated visual display for variable realizations, called a variation score, which enables the visual identification of variable clustering. A figure from Podesva (2008) is reproduced in Figure 3.4.3 for illustration.

![Variation score for “Heath” (from Podesva 2008)](image)

Figure 3.4.3. Variation score for “Heath” (from Podesva 2008)

In this presentation, modeled after musical score notation, each linguistic variable is arranged on a tier that extends over time with different possible displays for different
possible realizations. The variation score technique provides a deeper look into the relationship between sociolinguistic variables as they arise in discourse.

An older visualization technique, developed in the psycholinguistic tradition, that also provides visual insights into speech data at the level of discourse, is what Levelt (1989: 127) terms the Henderson graph after the technique’s creator (Henderson, Goldman-Eisler, and Skarbek 1966). The Henderson graph is a novel way to present and examine speech fluency. It plots phonation, or talk, time on the x-axis and pause time on the y-axis. An example, from Levelt (1989: 127), is provided here as Figure 3.4.4.

![Henderson Graph Example](image)

**Figure 3.4.4. A Henderson graph example (from Levelt 1989: 127)**

The Henderson graph provides not only a quick visual means to assess “fluent speech” versus “hesitant speech” (Levelt 1989: 127) – and through that, attention to speech – but
also a way to reconsider the analysis of speech fluency altogether. Through the visualization, in particular the overlaid slope lines (shown in Figure 3.4.4), we can conceptualize a new way to quantify notions such as *attention to speech* and paralinguistic cues.\(^\text{13}\)

\[3.4.4. \text{Generating visualizations through SLAAP’s transcript model}\]

The visualizations presented in the previous section provide new means to better understand our discourse or talk-in-action data. Yet, at first glance, they appear to have a methodological problem. Namely, how do you generate these visualizations? And, do the benefits of the visualizations make the work inherent in transforming the data justified?

The SLAAP transcript model solves this problem to a large degree. Since transcripts are finely time-aligned and transcript content are stored separately from their formatting, the SLAAP software can automatically produce visual displays of the transcript information. Additionally, by virtue of the timestamped and databased nature of variable tabulations within SLAAP, the realization of linguistic variables can be overlaid on these visualizations. In short, calling to mind the schematic of Figure 2.4.1, the various layers of metadata can be combined in powerful ways, enabling new insights into the overall data.

\(^\text{13}\) I discuss the *Henderson graph* further in §3.4.4 and at length in Chapter 10.
SLAAP currently allows for the automatic generation of two forms of “alternative” visualizations. Earlier, in Figure 2.3.1, (4) demonstrated graphicalization, one visual display of the unfolding of the speech event. Again, by virtue of the transcript model and its storage, users can instantly customize this graphical view. Figure 3.4.5 provides an excerpt of another graphicalization, here with overlaid variable tabulations, an audio player, and with the “ImageMap” feature enabled, which allows users to see the transcript text for a given line by placing the mouse over the graphical representation of an utterance.

Figure 3.4.5. Graphicalization screenshot from SLAAP (media: dca_keisha_a)

The graphicalization visualization does not provide as clear a view of variable realization as Podesva’s (2008) variation score. However, for examining the overall character of the discourse this presentation is, I argue, a useful way to explore speech and language variation. Further, SLAAP’s ability to automatically and instantly generate the graphicalization means it is always available as an option to SLAAP users. Once a transcript is in SLAAP, there is no added work required to explore the transcript through this – or any of the other – alternate formats.
SLAAP also allows for the generation of *Henderson graphs*. The original graphs presented by Henderson et al. (1966) were designed to examine temporal patterns in speech by single speakers.\(^{14}\) SLAAP seeks to advance the utility of this technique, by allowing for the plotting of multiple speakers and providing overlaid variable tabulation data. This is illustrated in Figure 3.4.6.

---

\(^{14}\) Henderson et al. (1966) examined both read speech and spontaneous speech. To examine spontaneous speech, however, they extracted uninterrupted responses to interviewers’ questions, and did not examine interactions between interlocutors, or representational techniques for dialogic data.
techniques can be readily implemented and automatically generated, and may provide us new, deeper insights into our data. Finally, it is not that there is something especially sophisticated about SLAAP’s software that makes this possible. Instead, it is the data model – centered on highly accurate time-alignment and databased storage – that readily lends itself to these transformations.

3.5. In closing: Transcription beyond text, annotation beyond transcription

The discussion in this chapter has focused on transcription, but readers will no doubt have noted that the underlying argument here extends beyond transcription to all sorts of linguistic annotation. SLAAP’s annotation and variable tabulation features are similar to transcripts in that they are units of annotation that are connected to the audio through timestamps.\(^\text{15}\)

In fact, scholars such as Bird and Liberman (e.g., 2001) have been arguing for some years now that all linguistic annotation types – and transcription is just one type of linguistic annotation – are fundamentally the same and that formalized methods should be implemented for the generation, querying, and maintenance of all of these annotation methods.\(^\text{16}\) The Bird and Liberman (2001) proposal is attractive because it formalizes the

\(^\text{15}\) The primary differences between transcripts in SLAAP and these other (meta)data types are (a) that transcripts must be generated outside of SLAAP (specifically in Praat) and then fed into the archive whereas annotations and tabulations are created ex nihilo within SLAAP; and, (b) transcripts’ data “points” have start and end timestamps while annotations and tabulations only have unitary timestamps.

\(^\text{16}\) I believe that Praat is so useful for transcription, in part, because transcription in Praat is treated no differently than any other type of annotation. Other transcription tools constrain the user by imposing their own (often theoretically-driven) transcription model on the user. Transcriber (Barras, Geoffrois, Wu, and
underlying argument of this chapter (and much of my work). While their formalization has not been implemented exactly within SLAAP, the data-model that has been implemented and discussed here is congruous with Bird and Liberman’s points, and supports their argument that all linguistic annotations are fundamentally similar, can be implemented in similar ways, and, most importantly, that there are benefits in doing so.

Liberman 2001), for example, implements a “turn” system that makes the accurate time-stamping of speaker overlap impossible. 

17 In fact, the Bird and Liberman (2001) proposal, the annotation graph, while underlyingly a data structural-model, is most conceptually accessible in a visual format, where in all cases labeled (annotated) arcs connect time-stamped nodes.
Part 2. Extending the variationist lens to speech rate and pause

4. Introduction to Part 2

4.1. Directions for extension: Investigating pause and speech rate

As discussed in the Part 1 of this dissertation, the sort of approach to language data instantiated by SLAAP enables the exploration of new sociolinguistic questions as well as new windows into traditional questions. This is particularly true of questions relating to sequential temporal patterns of talk – such as pause and speech rate – on account of the fine-grained time-aligned transcription method centered on the phonetic utterance. SLAAP allows large-scale (i.e., large N) corpus-like sociophonetic analyses of timing patterns through highly accurate, instrumental techniques. That is, with the tools developed in SLAAP, it is possible, for example, to extract for analysis tens of thousands of speech rate measurements from the archive almost instantly. In order to investigate some relatively “new” sociolinguistic questions, and to simultaneously exemplify the benefits of the SLAAP data and annotation model, Part 2 explores a few aspects of language and language variation by focusing on pause and speech rate.\(^1\) As I hope to

\(^1\) From the perspective of exemplifying some of the analytic benefits of SLAAP, these are really just two of the many features that could be examined here. SLAAP’s implementation of a time-aligned variable tabulation framework, for example, allows for the situating of particular instances of variables within the discourse context. I touch on this briefly in Chapter 10, but principally these benefits have been discussed
show, these analyses inform a range of issues in the larger sociolinguistic picture – such as the nature of the sociolinguistic variable and the possibilities of stylistic and register-based variation.

In fact, sequential temporal features, such as pause and speech rate, fall at the three-way nexus of sociolinguistics, psycholinguistics, and (formal) linguistic theory. Discoveries of socially conditioned, or correlated, aspects of speech rate and pause duration, for example, shed light on our understanding of the parameters of speech processing (Goldman-Eisler 1968), while the correlation between social aspects of speakers and tendencies of pause location in constituent structure might tell us something about social conditioning of, say, I-language (cf. Chomsky 1986). Finally, from a straightforward sociolinguistic variationist perspective, this study also has a number of bearings on “traditional” sociolinguistic variables and research questions, and the relationship between “paralinguistic” features and sociolinguistic style (qua Labov 1972).

I should point out that, while I believe the work of Part 2 speaks to this wide-range of theoretically rich questions within sociolinguistics and linguistics more broadly, my principal contributions here aim to be empirical and methodological. I necessarily leave some of the work of relating the foci, analyses, and findings of Part 2 to broader theoretical questions to the reader. Further, the purposes of the analyses here are not centered on showing specific regional or social patterns – I am interested in evaluating the degree to which pause and speech rate may vary along social parameters, such as

and illustrated in my broader work with Christine Mallinson on the corpus of interviews with African American girls in Washington, DC (e.g., Kendall, Mallinson, and Whitehead 2007; Mallinson and Kendall 2008, forthcoming).

2 This assumes, of course, that one believes in the idea that I-language can be socially conditioned.
region. Better understanding, for example, of the degrees to which pause distribution may vary between geographically proximate locations (such as different areas within North Carolina) is more interesting to me than questions focused on whether Southerners and Northerners or men and women or Blacks and Whites talk faster or slower than one another.

In the following chapters, we will explore various facets of pause and speech rate, moving from large-scale corpus-based quantitative analyses to finer-grained and case study-based examinations of pause and speech rate in attempts to inform our understanding of language processing, linguistic theory, and sociolinguistic variation.

4.2. The data

Since SLAAP’s transcript model (as described in Part 1) directly enables an accurate quantitative analysis of speech rate and pause, the data for these analyses come from many of the transcribed portions of interviews within the SLAAP archive. The transcripts within SLAAP are of widely varying lengths, ranging from a minute or so to over 70 minutes. For the macro-level analyses in Chapters 5 through 7, I have selected English language speech from speakers in as many transcripts as possible. I have excluded speakers from the analyses who appear in the transcripts with very limited talk.
in English. In later chapters (namely, Chapters 9 and 10) I draw on some data not included in the macro-level analyses.

Further, it also must be mentioned that for some speakers used in these analyses we have only a single transcript of five minutes from a single interview, while for others we have multiple transcripts spanning a number of interviews. The mean transcript length for all transcripts used in these analyses is 16.2 minutes. The shortest is 3.1 minutes long (though this transcript is for a speaker for which there is also another, longer transcript used). The longest transcript is 74.3 minutes. For Carissa, for example, the interviewer in the Washington, DC recordings with adolescent African Americans, we have over 12 hours of transcripts, with over 171 minutes of uttered talk – actual phonation – by Carissa (not including silent pauses). As these analyses progress, we will take advantage of these cases where we have a great deal of transcribed speech available in order to look deeper into the questions at hand.

More important than the length of each transcript is the amount of talk available for each of the speakers within a transcript. Especially in Chapters 5 through 7, speakers are treated independently from their interlocutors in their interviews and in some cases

---

3 The shortest transcripts in SLAAP are primarily from the “Dialect Bomb” set of recordings used for public demonstration or come from special recordings, such as the radio interview examined by Kendall and Wolfram (forthcoming). These very short excerpts are not used in the analyses.

4 Carissa is not used in the macro-level analyses of speech rate and pause (Chapters 5-7) mostly because she is the only speaker from Minnesota but also partly because the amount of data from her so heavily outweighs the available data from any other speaker. Her speech is used extensively, however, in Chapter 9, where I take advantage of this large amount of transcribed talk to better understand intra-speaker and interactionally-derived variation.

5 While the data come from SLAAP’s transcripts, we must understand them as coming from individual speakers contained in these transcripts. It would be incorrect to say that these transcripts are the data for the present analyses. Instead, I have tried to follow my own advice and keep in mind the thinking that the transcripts used here are interfaces to the data – not the data themselves.
only some of the speakers contained in a transcript have been selected for analysis. The decision of which speakers to include in the analysis and which to exclude was most often determined based on the amount of talk by the speakers. The speaker with the least amount of analyzed talk had only 17 utterances (i.e. stretches of talk surrounded by silence; cf. §3.2), but the median number of utterances across all speakers is 107, the mean is 229.5.

As is implied by my mention of Carissa, the interviewer, above, in addition to examining the interviewees in the SLAAP recordings I am occasionally examining the interviewers as well. As Kendall, Mallinson, and Whitehead (2007) argue, there are benefits to treating the interviewers in a recording as “speakers” for analysis. There are also many instances in the SLAAP archive where the interviewers are in fact locals of the research site and seem appropriate for inclusion. Some of the time the interviewers are as talkative as the interviewees. Also, often times, looking at the interviewers’ speech can tell us important things about the interviews as interactions (cf. Chapter 9).

Table 4.2.1 provides a summary of the demographic breakdown for the speakers used in Chapters 5 through 7, with respect to gender, ethnicity, and region. Readers may wonder what motivated the four-way division of North Carolina into Western NC, Central NC, Eastern NC, and Southern NC. In fact, my rationale for making this distinction was two-part. First, the data from the other regions, such as Ohio and South Texas, are not geographically disperse within those regions. So, for example, the Washington, DC speakers, all come from the same summer camp for teenagers of a specific background – and, in fact, we could readily describe those speakers as coming
from one community of practice (Eckert 2000) let alone a single geographic area (see §9.5, or Mallinson and Kendall forthcoming, for a fuller discussion of the Washington, DC interviews).

Table 4.2.1. Ethnicity, gender, and region of speakers used in analyses

<table>
<thead>
<tr>
<th>Ethnicity &amp; Gender</th>
<th>Ohio</th>
<th>South Texas</th>
<th>Western NC</th>
<th>Central NC</th>
<th>Southern NC</th>
<th>Eastern NC</th>
<th>Washington, DC</th>
<th>Ethn. &amp; Gender Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American Male</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>African American Female</td>
<td>1</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>European American Male</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>European American Female</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>5</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>Latino Male</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>Latino Female</td>
<td>-</td>
<td>10</td>
<td>2</td>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>Lumbee Male</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Lumbee Female</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Region Totals</td>
<td>7</td>
<td>24</td>
<td>8</td>
<td>29</td>
<td>19</td>
<td>5</td>
<td>12</td>
<td>104</td>
</tr>
</tbody>
</table>

The South Texas speakers also all come from the same research project conducted in a single fairly small town (Thomas and Ericson 2007; Wolford and Carter 2007). The Ohio data come from interviews distributed throughout the state, but six of the seven interviews examined come from northern Ohio, mostly suburbs of Cleveland; regardless, the limited number of speakers from Ohio made sub-dividing these speakers more precisely by region unrealistic.
The second, and primary, set of reasons for subdividing North Carolina into four regions comes from the fact that I have much larger amounts of data available to me from North Carolina, thanks to the work of the NCLLP. At the same time, North Carolina’s sociolinguistic history and geographic distribution of dialects makes, for example, lumping speakers from Eastern North Carolina with speakers from the Appalachian region of Western North Carolina obviously problematic (cf. Wolfram 1999). For these reasons I have separated the North Carolina speakers into four separate regions. Western NC includes speakers from the North Carolina mountain and foothill communities of Texana, Beech Bottom, and Hickory. Central NC includes speakers from central and northern North Carolina, including Raleigh, Durham, Princeville, and Warren County. Southern NC is entirely comprised of speakers from Robeson County. Eastern NC includes speakers from Hyde County, Roanoke Island, and Wilmington.

Age information is displayed in Figure 4.2.1, which displays these speakers by year of birth, organized by gender and ethnicity, with region indicated by the shape of the plotted points. The interviews examined here span 20 years of research, so it should be

---

6 It could be argued based on North Carolina’s sociohistorical development and dialect distribution that Warren County, in the north of the state, should be considered separately from the other locations deemed Central NC. This would be a reasonable argument. For the time being, I have defined the Central NC region in negative terms – not the Western mountains, not the Atlantic Coast, not the uniquely tri-ethnic Robeson County. I have also chosen to keep the Warren County speakers within the Central NC category to better balance the distribution of speakers. Future work will need to assess whether the Warren County speakers can be differentiated from the speakers to their south.

7 Readers may also wonder more broadly what motivated the use and definition of the “region” category whatsoever (the same could be said about the other social categories of ethnicity and gender). It would clearly be preferable, on the one hand, if there were enough data available from each specific community, to group speakers by specific community-location, for example, instead of the broader (and vaguer) “region” or, on the other, to consider these data more robustly in terms of criteria such as “cultural orientation”. As readers will see throughout these analyses, I am ultimately more interested in the possibilities of explaining the data by social categories. The social categories used here, like region, are intended as useful heuristics. Whether they prove to be epiphenomenal is less the point than discovering the extent to which sequential temporal features of talk pattern when examined from social vantage points.
noted that a plot by actual age at time of interview would show a slightly different
distribution than depicted in Figure 4.2.1. The colors and shapes of the plotting points
are used throughout Part 2 to indicate speaker gender, ethnicity, and region. The full
information for all of these speakers – including central tendencies for pause and speech
rate – can be found in the Appendix.

![Figure 4.2.1. All speakers plotted by year of birth](image)

In the following chapters, when I refer to speakers by name, I use the speakers’
identifications as they are given in SLAAP. Since these data come from diverse
sociolinguistic projects, a number of different speaker naming conventions have been

---

8 I somewhat arbitrarily display the information here by year of birth. In the later chapters, I analyze the
data by both age and year of birth. For the present purposes, neither “view” is better than the other and
including both here seemed superfluous.
used in the archive. Only in the case of the interviewers – sociolinguistic or sociological fieldworkers – will I provide a person’s real name. All other “names” are pseudonyms. Many speakers are referred to by initials and even these may be pseudonymous.

4.3. The tools

As was discussed in Part 1, one of the major features of the SLAAP software is the association of finely time-aligned transcript information to the audio files, in a dynamic and flexible way. SLAAP’s transcription method allows for the accurate capture of speech timing features, such as overlap and pause, since transcript lines are time-stamped to the audio and each line in a transcript corresponds to a phonetic utterance – that is, unbroken speech surrounded by silence on the part of the speaker. Pauses are accurately recorded as a matter of course as they are (time-stamped) blank lines in the transcript.

As was briefly indicated in Chapter 2 (and shown in (7) of Figure 2.2.1), I have developed a number of corpus-like features within SLAAP. These will be introduced and discussed further when relevant in the coming chapters. To extract the necessary data from the transcripts, I have written scripts in the statistical programming language R (R Development Core Team 2008), which communicate with the SLAAP server. These scripts can batch-process SLAAP’s corpus-like analysis features across all of the desired transcripts, combining the data from the many transcripts and preparing them (i.e. formatting them) for statistical analysis.
In addition to the analysis tools, SLAAP provides a number of interfaces with the transcript archive. To illustrate just one, Figure 4.3.1 shows a screenshot of the SLAAP transcript summary page. In this view, all of the transcripts (here, limited to a specific research project) are available along with information about the speakers in the transcript and the length of the transcript. Links are available to various transcript-based features.

![Figure 4.3.1. SLAAP screenshot of transcript summary for Pearsall, TX](image-url)
As an illustration of the sorts of information SLAAP can generate about each transcript, Figure 4.3.2 shows a screenshot from SLAAP’s transcription summary feature for the Washington, DC interview by Carissa with the girl “Elisa” (media: dca_elisa_a). This view gives us summary information about each transcript, including information about the total contributions to the talk by each participant.\(^9\)

---

\(^9\) “Talk-Time” is the measure of how much actual talk (phonation) is made by a given speaker. This measure does not include silent pauses. “Turn-Time” is the measure of how long a given speaker “holds the floor,” which includes intra-turn pauses. The amount of inter-speaker pauses and silence can be reconstructed by subtracting the total “Turn-Time” from the total length of a transcript. In other words, just over 14% of the transcribed interview consists of inter-speaker pauses, while 22% of the transcribed interview (2,316.94 sec – 1,728.15 sec) consists of intra-speaker pauses. So, overall, a third of the transcribed interview (actually, 36%) is silence.
4.4. Caveats for the present exploration

Before proceeding, a few caveats are in order. The data and analyses presented in Part 2 are – in a very intentional and literal sense – “explorations” of the data within SLAAP. They are not expected to provide definitive results. In particular, they take advantage of the cumulative data that are stored in the SLAAP archive. As is immediately seen in Table 4.2.1, above, these data are neither balanced nor representative in the senses that corpus linguists stress (cf. McEnery and Wilson 2001). At the time of this analysis, 130 transcripts are available in SLAAP and my work here is necessarily constrained by which interviews and segments of interviews have been transcribed by users of SLAAP. For example, despite the extensive work that the NCLLP has done in such communities as Ocracoke and Hyde County, there are no SLAAP transcripts for speakers from Ocracoke and only three for speakers from Hyde County. I have included here the three transcribed speakers from Hyde County, but obviously have not been able to consider Ocracoke in the present analyses.

Nonetheless, I have taken efforts to improve the representativity of the transcript collection. A number of the transcripts in SLAAP have been generated by NCLLP members and students in linguistics courses at Duke University and North Carolina State University,¹⁰ but I have created many of the transcripts myself with the present project in mind and in doing so have intentionally transcribed (segments of) interviews that flesh out social/demographic categories in the archive.

¹⁰ Although too numerous to mention by name here, I want to reiterate that I am deeply grateful to all of the individuals who have contributed transcripts to SLAAP.
An additional potential issue to address is the comparability of the transcripts in the SLAAP archive. In Chapter 3 I described some basic conventions for SLAAP transcripts and recently I wrote the first draft of the *SLAAP User Guide* (2008), which includes a long section describing transcription methodology for SLAAP. However, this sort of resource has not been available until fairly recently. I seek in the present project to make the best possible use of the data that have organically accumulated in SLAAP – that is, to support the claim I made in Chapter 2 that the accumulation of diverse language materials can be put to work to answer new questions that may not have been anticipated in the original research. A part of this is using the largest possible set of transcripts in SLAAP. This often involves comparing and analyzing transcripts from different transcribers and different research projects, and these sometimes have slightly different orthographic conventions.¹¹ For the sake of the present work, and the quantitative temporal explorations in the upcoming chapters, the differences in orthographic conventions are considered not to be a problem for the analyses I undertake. In the future, this will need to be more fully explored and any possible effects of the transcription conventions ameliorated. As discussed briefly in §3.2.2, it is possible that different transcribers have achieved different levels of accuracy in their temporal delimitation of the utterances (i.e., the accuracy of their time-stamping). However, I have

¹¹ For example, one major transcript contributor and I differ in the textual accuracy we aim for in our transcripts. This individual seeks to more accurately describe the actual productions of the speakers, whereas I typically aim to generate the most easily searchable transcripts through the consistent use of standardized spellings. An example of this is how a transcriber might choose to write a “mumbled” utterance of “I don’t know.” One could transcribe this as “I don’t know” capturing the lexemes and allowing for the easy retrieval of all occurrences in a transcript of “I don’t know” by a simple search for the string. Alternatively, one could transcribe the utterance closer to its actual production, as something like “I’on’ ow” which gives readers of the transcript more information about the actual production. Both approaches have their benefits and drawbacks.
either edited or transcribed in their entirety over two-thirds of the analyzed transcripts. I have also conducted exploratory comparisons of the transcripts’ temporal alignment. I do not believe that differences in transcriber accuracy are an issue for the present analysis, but this is likely an area for further consideration.

Finally, the data in SLAAP also, of course, come from sociolinguistic interviews and are not always the best sort of data for all of the explorations I would be most interested in. This is especially true in the three chapters immediately following this one, where I explore whether pause duration and speech rate correlate with social attributes of speakers. As readers will see, in Chapters 8 and 9, I quickly have to acknowledge that style and register are important factors in pause and speech rate. More complete “answers” to the questions I pose in these chapters may have to wait for more finely controlled recordings with higher quality audio on the one hand, such as lab recordings,12 and actual natural spoken conversation on the other, such as the sort of data – recordings of spontaneous talk over the course of a day – reported on in Hindle (1980) instead of interview talk.

---

12 Some recent papers by Clopper, Smiljanic, and colleagues (Clopper and Smiljanic 2007, Armstrong, Clopper, and Smiljanic 2008) have investigated variation in pause duration and speech rate using laboratory read passages of texts. While this might appear to mitigate to a large degree the effects of stylistic and interactional differences, it actually may increase stylistic differences in the data. Some speakers may adopt a slow, pedantic “read” style, while others may read the texts quickly, with less “literary effect” (cf. Kowal and O’Connell 1980). In short, it is not at all clear that laboratory speech minimizes stylistic variation or provides a controlled setting for examining temporal features of speech. (This important point was made by an audience member at Armstrong et al.’s 2008 presentation – unfortunately, I do not remember who exactly it was that deserves the credit for this observation.)
4.5. *In closing: A corpus-like approach*

In closing this introductory chapter I point out that I seek to approach these data in a large-scale, corpus-like way in the following chapters. I hope for this analysis to serve as a proof-of-concept as much as I hope for it to illuminate areas of language variation previously little examined. Clearly, the analyses presented in the following chapters are putative to lesser or greater degrees; they need to remain contextualized within the caveats that have been presented here. In short, there is much more work to be done than what is contained here, so, more than anything else, it is hoped that these chapters justify and inspire this needed future work.
5. Toward a quantitative sociolinguistic analysis of pause

5.1. Introduction

Silence in speech is a critical part of expression. A large proportion of talk in action is, in fact, silence – that is, comprised of the pauses between speakers’ utterances. Sociolinguistics has often understood language as situated practice within social and socialized contexts (Eckert 2005; but see even as far back as Sapir 1921, cited in Chambers 2003), and as such, we would – I think – expect that pause realization might correlate with social characteristics of speakers, especially since silences make up so much of actual conversational talk. Nonetheless, pause has received only occasional focus in linguistics, and then this focus has primarily come in the domain of psycholinguistics, where for a number of decades beginning in the 1950s pause was considered of import for understanding language processing.

Some of the findings from within psycholinguistics were extremely provocative from the perspective of language variation studies. Frieda Goldman-Eisler, a prominent psycholinguist and pioneer of pause studies, for example, described some of her findings thus:

Pausing during the act of generating spontaneous speech is a highly variable phenomenon which is symptomatic of individual differences,

---

1 See Figure 4.3.2 and footnote 9 of Chapter 4 for an example of this from SLAAP. In that transcript, we see that over 1/3 of the transcribed interaction is comprised of silence on the part of both speakers.
sensitive to the pressure of social interaction and to the requirements of verbal tasks and diminishing with learning, i.e. with the reduction in the spontaneity of the process (Goldman-Eisler 1968: 15).

Beyond this noting of “individual differences” and sensitivity to “social interaction,” however, the question of whether there are differences in pause realization among different groups, however socially defined, has surprisingly been rarely pursued.

In fact, within sociolinguistics pause has been analyzed less often and mostly only for discourse analytic purposes (most notably, perhaps, in Mendoza-Denton’s 1995 analysis of the Anita Hill-Clarence Thomas hearings) or qualitatively in terms of intercultural communication (e.g., Philips 1976; Lehtonen and Sajavaara 1985).

Traditional variationist thinking seems to have taken the stance that pause cannot be studied linguistically. Ronald Macaulay strongly espouses this view in his essay on “Discourse Variation” in the widely read *Handbook of Language Variation and Change* (2002).

One of the most common functions of discourse is to communicate something, but the proper study of linguistics is not communication. (In this case I agree with Chomsky.) Linguists are concerned with the use of language in communication, but that is a very different thing. To take an obvious example, conversation analysts … and psychologists … have shown the significance of pauses and silence in communicating. However, *there can be no linguistic analysis of silence, though pauses may be a guide to linguistic units.* (Macaulay 2002: 284, emphasis added)

This declaration seems to me strikingly reminiscent of earlier thinking about language variation in general – that most variation is “free” variation, and outside the possibilities for systematic analysis (cf. discussions in Labov 1966; Chambers 2003).
While many sociolinguists would not necessarily agree with Macaulay’s statement – at least that boldly – it does indicate an implicit position that the majority of variationists have long accepted: there are certain things that we study and certain things that we do not.\(^2\) This could probably be said about the history of linguistics more broadly, but it is especially true of the variationist paradigm, where the quantitative focus has often been on a discrete set of phonological or morphosyntactic variables. With the growth of sociophonetics and variationists’ recent turn to analyzing prosodic variation (e.g., Thomas and Carter 2006; Clopper and Smiljanic 2007; Vaughn 2008), it is a ripe time to turn our thorough attention to possible social factors behind pause production.

Contrary to Macaulay’s remark against the possibility of conducting a linguistic analysis of pause, this chapter seeks to add a richer dimension to the literature on pause by approaching *silent pauses*\(^3\) from a variationist, quantitative perspective. The development of quantitative approaches to studying variation in speech timing opens up both an array of new *linguistic variables* for sociolinguistic description (cf. Wolfram 2006) and a range of new windows into how sociolinguistic variation can be understood in interaction (cf. Mallinson and Kendall 2008) and in relation to language processing (Kendall 2008b). I believe that the examination of variation in pause may provide new insights into linguistic theory – the ultimate goal of variationist linguistics (Labov 1966).

\(^2\) Interestingly, in his 1991 book, Macaulay argues for expanding variationist focus beyond “a few variables” (5). See §1.3.1 for a fuller quotation. I have trouble reconciling his two apparently opposing stances.

\(^3\) With the exception of a brief consideration in §5.5, I am only discussing *silent pauses* and not what are normally referred to in the literature on pause as *filled pauses*. Filled pauses are hesitation (or discourse) markers like “um” or “uh” – there are some interesting papers on filled pauses in the psycholinguistic tradition (e.g., Clark and Fox Tree 2002), but I will be focusing only on silent pauses here, even when I say only “pause”.

93
I begin by providing a short literature review of linguistic approaches to pause. I then undertake a quantitative analysis of variation in the intra-turn silent pause durations of the speakers outlined in §4.2 (and enumerated in the Appendix) and explore correlations in pause duration with social attributes of the speakers (ethnicity, gender, age, and region). I will end this discussion with a brief deeper examination of a subset of these data, examining pause durations in terms of a typology of pause and the co-occurrence of filled pauses.

5.2. Research on pause

As indicated by the quotation from Goldman-Eisler (1968) above, in the psycholinguistic literature, pause has primarily been considered to be an outcome of processing activity. For example, Goldman-Eisler (e.g., 1958, 1968) showed that much of spontaneous speech is “a highly fragmented and discontinuous activity” (1968: 31), that pauses are more likely and longer before words with less predictability and with more difficult speaking tasks, and that – in the terminology and conception of the time – pauses can be used “to sort out which parts of verbal sequences are verbal habits and which are being created at the time of speaking” (1968: 43). Additionally, Goldman-
Eisler found that pauses account for much of the variation in perceived speech rate, a finding that we will return to in the next chapter.4

Goldman-Eisler’s work is paralleled by the findings from other psycholinguists who have pursued questions of speech timing. For example, Maclay and Osgood’s early (1959) work “famously” (according to Deese 1980: 96) found that hesitation pauses are more often realized before a semantically heavy unit than at clause boundaries.5 In general, Goldman-Eisler’s various findings appear to have been confirmed numerous times and in numerous ways (e.g., Kircher, Brammer, Levelt, Bartels, and McGuire 2004; see, more generally, Levelt 1989).

An interesting example of some related research is found in Wallace Chafe’s work in the *Pear Stories* (Chafe 1980a), in which Chafe used pause to help better understand the unfolding of information flow in discourse.6 In particular, he views “hesitation phenomena … as overt, measurable indications of processing activity” (Chafe 1985: 78) and examines correlations between pause realizations and “foci of consciousness” (“ideas” in his 1980a terminology) in speakers’ recollections of a previously viewed film. He focuses on how pause location and duration relate to the cognitive tasks of speakers’ determination of *what* to talk about and *how* to talk about it. While Chafe does not focus in depth on a quantitative analysis, he finds that a higher proportion of pauses fall between “focus clusters” than fall within them, and that the

---

4 This is the primary reason this chapter comes before the chapter on speech rate! See the following chapter, especially §6.2, on the different conceptions of “speech rate”, some are pause inclusive while others (including the measure adopted in this study) are pause exclusive.

5 However, note that Kowal and O’Connell (1980) found conflicting results with respect to pause location – that pauses are more frequently aligned with function words and not content words – and argue that Maclay and Osgood’s (1959) methodology was problematic.

6 Chafe’s work here is, of course, related to his work on intonation units, which was discussed in §3.2.
pauses occurring between clusters have a longer mean duration than those within clusters (Chafe 1985).

Daniel O’Connell and Sabine Kowal have a long history of interest in “pausological research” (cf. Kowal and O’Connell 1980). They credit the main hypothesis of their line of research directly to Goldman-Eisler, stating that one can map “a lawful relationship between temporal phenomena in human speech and concurrent cognitive processes” (Kowal and O’Connell 1980: 61). O’Connell, Kowal, and colleagues looked at pause length and frequency as a function of age and language learning (cf. O’Connell and Kowal 1972; Kowal, O’Connell, and Sabin 1975). They reported:

We have tentatively associated the length of silent pauses with the generation of meaning or a more cognitive aspect of processing, whereas we feel that frequency of silent pauses reflects structural aspects or linguistic execution of semantic planning. In any event, younger children are unable to think and talk at the same time (Kowal and O’Connell 1980: 63).

They find adults, on the other hand, to have a “remarkable stability in speech rate and silent pause usage” (1980: 63).

Occasionally, O’Connell, Kowal, and colleagues have taken interest in broader social factors in pause and speech rate beyond foci on cross-linguistic comparisons and age-grading (e.g., O’Connell and Kowal 1972). They report some consistent differences between genders in experiments with younger speakers, finding that boys tend to have longer and more pauses than girls in out-loud reading and narrative production (Kowal and O’Connell 1980: 66; Kowal, O’Connell, and Sabin 1975). They have found some
evidence that young children in urban lower socioeconomic situations have longer pauses than their peers in other situations, but that by second grade the differences were eliminated (Bassett, O’Connell, and Monahan 1977). All in all, these experimental studies have generated provocative, though putative, findings about socially-based variation in pause production but it does not appear that they have been followed up on to any depth in the following decades.

S. R. Rochester’s (1973) article titled “The Significance of Pauses in Spontaneous Speech” provides an excellent early review of pause work beyond the projects of the scholars mentioned above. In addition to focusing on psycholinguistic models of the speaker and how silent and filled pauses may serve as clues to the process of speech production, Rochester also reviews “the function of pauses for the speaker” (1973: 65) in the psycholinguistic literature, which he describes as focusing on questions of cognitive load (i.e. “task difficulty”) and affective state (i.e. “anxiety”). As we might expect, most of the studies reviewed by Rochester consider the speaker “simply as a language generator which pauses either in the course of normal decision-making operations or because of disruptions in those operations” (1973: 74). However, he also discusses a handful of studies that approach pause from a more social psychological perspective. Some of the relevant findings presented by Rochester for a sociolinguistic consideration of pause include the following: “Subjects scoring high in an audience sensitivity test paused more frequently when addressing an audience than did low scorers” (Rochester 1973: 75); and, “pause frequency remained constant but duration increased when
utterances of subjects scoring high in concern for approval … and extroversion … were compared with the vocalizations of low-scoring subjects” (Rochester 1973: 75).

Beginning in the late 1970s, Stanley Feldstein and his colleagues undertook a number of related projects, examining “conversation chronography,” the timing of speech sounds and silences and the role that these timings have on “the impressions that interactants form of one another” (Crown and Feldstein 1985: 32). Their examinations ranged from inquiries into the level of accommodation between interlocutors (Crown and Feldstein 1981, discussed in Crown and Feldstein 1985) to the relationship between actual speech production and the stereotyped notions of speech timing by extroverts and introverts (Feldstein and Sloan 1984). Importantly, a number of Feldstein’s experimental findings point to the formation of different impressions by hearers on aspects of pause depending on social attributes of the speakers, such as ethnicity (Feldstein and Crown 1978, discussed in Crown and Feldstein 1985) and gender (Feldstein and Crown 1978, discussed in Crown and Feldstein 1985; Feldstein, Dohm, and Crown 1993). In sum, they found “the perceptions of the conversationalists were complexly related to the temporal patterns of their verbal exchanges primarily as a function of their race and gender” (Crown and Feldstein 1985: 42). In other words, gender and ethnicity appear to interact with speech timing features in influencing speakers’ perceptions of one another. So, while it is clear that pause has a cognitive, psycholinguistic component from the earlier work in psycholinguistics (such as by Goldman-Eisler 1958, 1968; Maclay and Osgood 1959; etc.), it is also clear that pause has a social component outside of being the outcome solely of mental processes.
At the same time – despite the paucity of explicit investigations in the literature into this further – this “social component” can be seen in terms of pause production when we look at cross-cultural differences in the communicative use of silence and pause. For example, we see this qualitatively when we compare many of the contributions in Tannen and Saville-Troike’s (1985) volume, *Perspectives on Silence*. Tannen’s (1985) New York Jewish Conversational Style, with its avoidance and negative view of silence, contrasts starkly against “The Silent Finn” of Lehtonen and Sajavaara (1985; Sajavaara and Lehtonen 1997).

At a purely quantitative level, Campione and Véronis (2002) compared pause duration across five European languages (English, French, German, Italian, and Spanish) by analyzing approximately 6,000 pauses in about 5½ hours of recorded speech. They found that there are differences in pause length between languages (in particular, Spanish had a median pause duration of about 100 ms longer than the other languages – 587 ms vs. ~ 490 ms).

As Saville-Troike (1985) tells us, “within linguistics silence has traditionally been ignored except for its boundary-marking function, delimiting the beginning and end of utterances” (3). From a corpus linguistics perspective, especially, this focus on pause as a delimiter of speech is not surprising since, at the most basic level, pause serves to separate strings of speech from one another (cf. Mukherjee 2000).\(^7\) Pause has also played a similar boundary-marking role in variationist linguistics in that for some variables it has

\(^7\) And, of course, as discussed in previous chapters (and Kendall 2006-2007, 2007a), pause plays a primary role in the determination of what constitutes an “utterance” (a transcript line) within SLAAP’s transcription system.
been found to be a significant constraint. The major example of this is syllable-coda 
consonant cluster reduction (CCR), or t/d deletion, where numerous studies (e.g., Guy 
1980; Wolfram, Childs, and Torbert 2000) have found following pause to constrain 
consonant cluster reduction differently than following consonant or vowel environments. 

As I mentioned above, sociolinguists have recently become interested in 
understanding prosodic variation and a small movement has begun towards investigating 
questions around pause (and speech rate, which is addressed in the next chapter). In a 
2006 conference paper, I asked whether pause could be considered a sociolinguistic 
variable (in the sense of Wolfram 1993) and found favorable results. Recently, Clopper 
and Smiljanic (Clopper and Smiljanic 2007; Armstrong, Clopper, and Smiljanic 2008) 
have investigated regional and gender variation in pause (and speech rate) and asked 
whether pause duration was a factor in the stereotype that Southerners talk slower than 
Northerners.

So, while the linguistic literature on pause is relatively small, it is broad, ranging 
from discourse analytic and qualitative to psycholinguistic or corpus-based and 
quantitative. What’s missing here, and what motivates the present study, is an explicit 
investigation of the relationship between pause production and social meaning at a more 
u nuanced quantitative level than between geographically distant and culturally separate 
populations. That is, do groups (however socially defined) perform or index (or whatever 
we want to call it) their group identity through their pause practices? Further, is the 
variation we find in pause realization – even such as that found by Campione and Véronis 
(2002) in their comparison of European languages – linguistically and socially
meaningful? Wolfram (2006) reminds us “the empirical reality is that the boundaries of significant and insignificant language variation are often gradient and obscure rather than discrete and transparent” (334). Does variation in pause fall within the realm of significant language variation? The SLAAP software and data-model provide us the tools to assess this question.

5.3. Defining and measuring silent pause

In order to conduct an analysis of pause, I have developed a pause analysis feature in the SLAAP software that allows for the automatic retrieval of information about speakers’ pauses in the transcripts in which they appear. This feature is illustrated by the screenshot in Figure 5.3.1. While SLAAP extracts all of the relevant pauses for a speaker, compiles these pauses in a “pause sheet,” and provides summary statistics for that speaker, it does not (yet) on its own have an easy mechanism for comparing speakers or conducting more sophisticated statistics on the data. To automate the process – that is, to conduct a large-scale corpus-like study – I have further written a set of scripts in R (R Development Core Team 2008) that interact with SLAAP and download all of the “pause sheets” to my local computer. These scripts then allow me to conduct sophisticated statistical analyses with relative ease, both with R’s built in functionality and with the Rbrul package (Johnson 2008, 2009).
The extraction process is a fairly straightforward task, since the transcripts are accurately coded for pause automatically in the transcription process. However there are a number of methodological issues that warrant attention. I review these next.

Figure 5.3.1. Screenshot from SLAAP showing an analysis of pause

5.3.1. Count versus don’t count pauses
One of the primary difficulties in the analysis of pause is the determination of *count* and *don’t count* forms. This is in general no easy task for any sociolinguistic variable. For example, Blake’s (1997) discussion of don’t count forms for the analysis of the copula in African American English shows just how difficult this can be, even for more traditional variables. Pauses – as researchers often note – appear most commonly at turn boundaries. As Norma Mendoza-Denton (1995) has shown through her analysis of *gap-length* (the silence between the speech of two speakers) in the Anita Hill-Clarence Thomas hearings, the pauses occurring between turns may have dramatic effects on the interaction as a whole. However, in order to analyze pause as a linguistic variable for individual speakers, only those pauses that occur turn-internally (and without interruption by an interlocutor) can be counted – otherwise we have no way of knowing to whom to attribute the pause. In other words, counted pauses must occur within an uninterrupted turn by the same speaker. Turn-external pauses seem to be a variable of importance at the interaction- or discourse-level, but not of use for investigating variation at the level of the individual speaker, the focus of the present endeavor.

5.3.2. Measuring pauses

Even accepting an exclusive view on only turn-internal pauses, there remain two important quantitative methodological questions. The first is how much silence it takes to qualify as a pause. The second is the adoption of a central tendency measurement that is most reliable when measuring pause. I address these here in order.
As to the first question, measurement *thresholds* seem to be the norm in analyzing pause. Having a low and, possibly, a high cut-off is necessary so that pauses of a certain shortness or length are excluded from measurement. Kowal and O’Connell, for example, adopted “as a convention the minimal cut-off point of 270 milliseconds for silent pauses” (Kowal and O’Connell 1980: 62), while Goldman-Eisler and colleagues adopted various low threshold values from 100 milliseconds (e.g., Henderson, Goldman-Eisler, and Skarbek 1966: 208) to 250 milliseconds (e.g., Goldman-Eisler 1968: 12), depending on the experiment.

In their quantitative analysis of pause duration differences between European languages, Campione and Véronis (2002) explain that – despite the fact that “silent pauses shorter than 200 ms are very difficult to discriminate from occlusives and taking them into account requires enormous manual effort” (202) – using thresholds skews the resultant data. However, for this research using some threshold value appears necessary, since transcribers vary in terms of how small a boundary they choose (or are able) to delimit. The recordings used here also come from field interviews, which have varying degrees of background noise, and this can make the detection of very short pauses difficult and unreliable. At the same time, since “enormous manual effort” goes into the SLAAP transcription process in the first place – and modern computer-based tools (namely Praat, Boersma and Weenink 2007) allow for exceptionally fine-grained temporal accuracy – my experience has been that 200 ms is a much higher threshold than necessary, and that pauses on the order of 60 ms are reliably captured so that number is
used here for the analysis.\textsuperscript{8} At the upper end of the range of realized pause lengths, a high threshold of 5 seconds has been used, even though intra-turn pauses longer than 2 seconds are rare in the data and exploratory analyses have indicated that the use of a high threshold may not be necessary at all.\textsuperscript{9}

Campione and Véronis (2002) also address the second measurement question, which central tendency measurement is most useful for the analysis of pause. This is especially important since pause duration shows a wide range of variance. They find that “arithmetic mean is not a reliable measure of central tendency” (200) for pause. On the other hand, using the median value for pause provides a rather stable measurement of the central tendency, since it has the advantage of filtering out those occasionally outlying measurements and helps to compensate for any potential effects from the choice of threshold values. In the following analysis, I typically provide median, mean, and standard deviation measures, though I suggest readers pay most attention to the median values.

5.3.3. Hesitation pauses v. respiratory pauses

\textsuperscript{8} A large number of intra-turn pauses are shorter than 200 ms in duration (23\%, or 5,123 of the 22,734 pauses examined here were between 60 ms and 200 ms in duration), a fact that makes this decision an important one. In my future analyses, I plan to investigate the impact on the quantitative analysis of distinguishing between short (60 ms to about 250 ms) pauses and longer pauses.

\textsuperscript{9} In fact, after thresholding – i.e. retrieving all, and only, pauses between 60 ms and 5 seconds – only 2.7\% (623) of the total (22,734) counted pause measurements were above 2 seconds in duration.
There remains one final area to discuss before proceeding with the analysis: Namely, the fact that all silent pauses are not equal. As Goldman-Eisler (cf. 1968: 12) first enumerated, we can readily distinguish three types of discontinuities in the speech signal. These are:

(a) Discontinuities in the speech signal related to articulatory phenomena, such as the brief gaps between stop consonant articulations (e.g. fully articulated take care or, conveniently, stop gap);

(b) Discontinuities related to the inhalation or exhalation of breath; and,

(c) Discontinuities unrelated to articulation, (a), and respiration, (b).

Category (c), then, includes all pauses that are not the result purely of physiological processes, in other words, pauses that are cognitively, pragmatically, or discourse-related. Ideally, we are interested only in those pauses that fall under category (c). Following Goldman-Eisler and her colleagues, we can label this category (c) as hesitation pauses.

As outlined above, we use a low threshold to remove the articulatory phenomena, (a). They are, as a rule, quite short (< 60 ms), so can easily be discounted. The question of how to differentiate between respiratory pauses and non-respiratory (i.e. hesitation) pauses, however, is a harder one. Deese (1980), for example, discusses the possibilities of and problems in differentiating actual hesitation pauses from intentional or communicative pauses. He notes the need to examine speech “in its complete context” (72) and that often this can only be done – if at all – with video records of the speech event.
Some laboratory and experimental work – such as that undertaken by Goldman-Eisler (e.g., 1968) and some of the other researchers discussed in §5.2 – has been in a position to record or account for breathing, either through special recording equipment or just by way of having a high-enough quality acoustic environment that breath is audible in the recording.\textsuperscript{10} Unfortunately, many of the sociolinguistic field recordings in SLAAP and used here are too noisy or otherwise not of high enough fidelity to allow breathing to be coded with necessary reliability.

For the current study, I do not pursue the differentiation of hesitation pauses and pauses that are purely for breath. Goldman-Eisler (1955, 1968) examined the relationship between breath pauses and hesitation pauses and found that

breathing might normally occupy between 1.5 and 15 seconds in a minute, i.e. between 2.5 and 25\% of the total speaking time. Measurements of periods of hesitation based on the same samples on the other hand showed these to occupy an average of between 40 and 50\% of the total speaking time of one person (Goldman-Eisler 1968: 24).

In other words, hesitation pauses account for the larger proportion of discontinuities and, further, breathing appears to be “a passive process fitting into given breaks in speech irrespective of whether or not these occur at grammatical junctures, and that the decisive factor in breaking up the linguistic groupings at non-grammatical places is hesitation” (Goldman-Eisler 1968: 98).

Additionally, some of Grosjean’s research (e.g. 1980a; Grosjean, Grosjean, and Lane 1979) investigated the question of whether breath pauses and non-breathe pauses

\textsuperscript{10} Goldman-Eisler (1955, 1968: 96) reports that locating breathing is relatively straightforward provided that the recordings are of high enough quality.
overlap and found that, in terms of pause location, that “there were no systematic differences between the breathing and no-breathing conditions” (Grosjean 1980a: 97). Bolstered by Goldman-Eisler’s and Grosjean’s findings, I have not attempted to differentiate between these two categories of pauses, hesitation versus breath. All things being equal, it also seems likely that, across speakers, breath pauses – with a physiological function – are more regular overall (as they are linked to universals of human physiology) than hesitation pauses. If this is the case, lumping breath pauses into a study of hesitation pauses probably only decreases the noticeable meaningful variability (i.e. would lead to Type II errors), which is a better analytic outcome than if it might lead to larger patterns of variation that are not meaningful or real (i.e. Type I errors). At the same time, it is clear that categories (b) and (c) often overlap – how does an analyst differentiate breath pauses that co-occur with language planning or pragmatic function and those that are merely a break for breath? Intuitively, it does not seem likely that many breath pauses occur with the absences of planning or other overt hesitation. In fact, conversely, speakers, at least to a certain degree, likely coordinate their breathing with their language planning processes, as Grosjean’s (1980a; Grosjean, Grosjean, and Lane 1979) work appears to indicate.

In sum, we move ahead with this analysis feeling confident that we have excluded articulatory, or type (a), pauses from the analysis, and realizing that, although not ideal, the lumping of breath pauses, type (b), with hesitation pauses, type (c), is a necessary maneuver, the implications of which are assumed to be small.
5.4. Analysis and discussion

We turn now to the analysis of the overall pause duration data from the 104 speakers introduced in §4.2. There are 22,734 pause measurements in the dataset after thresholding. Examining all of the pauses, it became clear that pause has an approximately log-normal distribution\(^{11}\). The distribution of the pause data is displayed in Figure 5.4.1, in log-ms. Most of the data (19,365 of the 22,734 pauses, or 85.2\%) are below 1.0 second in duration (6.91 log-ms). The mean duration of all the pauses is 574 ms (6.35 log-ms),\(^{12}\) with a standard deviation of 544 ms. The median pause duration is 412 ms (6.02 log-ms).

![Figure 5.4.1. Histogram of log-pause durations](image)

---

11 In all cases, the log function used in this work is the natural log, \(\ln\).
12 The log-mean, i.e. the mean of the log-pause values, is 5.99 log-ms.
Figure 5.4.2. Pause duration by speaker, ordered by median pause and displayed in milliseconds in logarithmic scale.
Figure 5.4.2 shows boxplots (Tukey 1977; Benjamini 1988) for all speakers, ordered by median pause duration and displayed on a logarithmic scale. Color (although not available to all readers) is used to indicate each speaker’s gender and ethnicity. Dark green is used to display white males, light green for white females, dark blue for black males, light blue for black females, light brown for Latino males, yellow for Latina females, red for Lumbee males, and pink for Lumbee females. The mean value across all speakers’ median pause durations is 401 ms.

The Appendix lists the complete data for all 104 speakers used in the analysis for this and the following two chapters. Tables 5.4.1 – 5.4.3 summarize these speakers’ pause data for ethnicity, gender, and region, respectively. The rows of each table are ordered by median duration “by measurement”.

<table>
<thead>
<tr>
<th></th>
<th>By Measurement</th>
<th>By Speaker (based on speaker’s median pause dur)$^{13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of tokens</td>
<td>Median</td>
</tr>
<tr>
<td>African American</td>
<td>12,747</td>
<td>422 ms</td>
</tr>
<tr>
<td>European American</td>
<td>3,930</td>
<td>409 ms</td>
</tr>
<tr>
<td>Latino</td>
<td>5,463</td>
<td>394 ms</td>
</tr>
<tr>
<td>Lumbee</td>
<td>594</td>
<td>367 ms</td>
</tr>
<tr>
<td>Overall</td>
<td>22,734</td>
<td>412 ms</td>
</tr>
</tbody>
</table>

$^{13}$ That is, the “By Speaker” fields in Tables 5.4.1 – 5.4.3 (and next chapter’s Tables 6.5.1 – 6.5.3) display the group median, mean, and standard deviations calculated from the individual speakers’ median values. So, for example, there are 34 African American speakers and the median value of those 34 speakers’ individual median values for pause duration is 376 ms.
Table 5.4.2. Pause data by gender

<table>
<thead>
<tr>
<th></th>
<th>By Measurement</th>
<th>By Speaker (based on speaker’s median pause dur)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of tokens</td>
<td>Median</td>
</tr>
<tr>
<td>Male</td>
<td>6,904</td>
<td>427 ms</td>
</tr>
<tr>
<td>Female</td>
<td>15,830</td>
<td>404 ms</td>
</tr>
<tr>
<td>Overall</td>
<td>22,734</td>
<td>412 ms</td>
</tr>
</tbody>
</table>

Table 5.4.3. Pause data by region

<table>
<thead>
<tr>
<th></th>
<th>By Measurement</th>
<th>By Speaker (based on speaker’s median pause dur)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of tokens</td>
<td>Median</td>
</tr>
<tr>
<td>Eastern NC</td>
<td>658</td>
<td>456 ms</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>9,495</td>
<td>451 ms</td>
</tr>
<tr>
<td>South Texas</td>
<td>3,715</td>
<td>425 ms</td>
</tr>
<tr>
<td>Western NC</td>
<td>1,528</td>
<td>390 ms</td>
</tr>
<tr>
<td>Central NC</td>
<td>5,299</td>
<td>375 ms</td>
</tr>
<tr>
<td>Southern NC</td>
<td>1,261</td>
<td>361 ms</td>
</tr>
<tr>
<td>Ohio</td>
<td>778</td>
<td>302 ms</td>
</tr>
<tr>
<td>Overall</td>
<td>22,734</td>
<td>412 ms</td>
</tr>
</tbody>
</table>

Using the Rbrul statistical package (Johnson 2008, 2009), I submitted the entire matrix of pause data (22,734 tokens) to a mixed-effect model analysis testing the effect of ethnicity, gender, region, age, and year of birth on pause duration measurement.

Mixed-effect model analysis is a recent addition to the statistical “toolbox” of sociolinguistics. As Johnson (2009) argues, it has many advantages over the near-ubiquitous Varbrul-type logistical regression implemented in many popular sociolinguistic statistical analysis packages (such as GoldVarb; Sankoff, Tagliamonte, and Smith 2005). Johnson’s Rbrul package extends the basic logistic regression features of GoldVarb, but also supports continuous variables. In addition to this ability to
incorporate continuous variables as predictors and dependent variables, an obvious necessity here, Rbrul’s support for mixed-effect model analysis is especially useful for the data here, because the method is designed to work with multilevel data. That is, mixed models allow the incorporation of individual speakers as grouping, or random, effects – an approach that builds a much more robust and honest statistical model (cf., Quené 2008; Johnson 2009). Instead of treating each token as an independent measurement, random effects capture the fact that some predictor variables, such as individual speaker, can have idiosyncratic tendencies that obscure the overall patterning of the data.

For the model here, speaker is included as a random effect. This approach also mitigates the fact that each speaker contributed fairly widely different Ns to the overall data – that is, that the data are unbalanced in terms of individuals’ contributions.14 Additionally, this helps to account for the potential that pause durations (and speech rates) may be highly idiosyncratic (as indicated by Goldman-Eisler’s, e.g. 1968, work) at the same time as they may be influenced by social categories (such as ethnicity and gender).

Importantly, Rbrul’s output – especially as presented here – is different from the more familiar Varbrul-type output in that it does not report the results in terms of factor weights. Instead it presents the strength of effects for specific factors in terms of log-odds (or logits). Log-odds are literally log-transformations of the odds of an outcome (cf.

14 The median number of tokens (i.e. pause duration measurements) per speaker is 104, but there are half a dozen speakers or so with token counts as high as 1,000.
Positive values indicate favoring effects and negative values indicate inhibiting effects. Log-odds have two primary benefits over factor weights: They are additive—that is, predicted outcomes of a model can be understood in terms of adding all of the log-odds for the relevant factors of the independent variables—and, perhaps more importantly for the work here, factor weights are not meaningful for continuous variables—log-odds are necessary for understanding the effects of continuous predictors and continuous response variables.

The results of the mixed model analysis for the pause data are displayed in Table 5.4.4. Although pause distributes log-normally, it is important to note that the pause data submitted to the mixed model analysis are the raw pause measurements. This was determined not to affect the outcome of the model in terms of which factor groups are found to be significant or in terms of the ordering of the factors within each factor group, and it only marginally impacts the actual $p$ values. Analyzing the raw, non-log pause data makes interpreting the results more straightforward, so I have opted for this approach.

This model selects region ($p < 0.0001$), gender ($p < 0.0001$), and ethnicity ($p < 0.01$) as significant influences of pause duration.

---

15 The odds of an outcome are calculated as $p/(1-p)$, where $p$ is the probability of the outcome. So, the log-odds of an outcome are $\log(p/(1-p))$.

16 For sake of space, and because there are other sources that detail information about mixed-effect models, I do not provide here a fuller explanation of mixed-effect modeling. Pinheiro and Bates (2000) is a good textbook on mixed-effects modeling in general, with an emphasis on programming in the statistical languages S and R. Johnson (2007, forthcoming) thoroughly discusses the use of mixed-effect models specifically for sociolinguistic analysis (see also Quené 2008), and Baayen (2008) provides a good introduction to statistics in R for linguists, including a section on mixed models. The reader is referred to these sources for more information about mixed-effect modeling.
### Table 5.4.4. Mixed-effect model for all pause duration data

<table>
<thead>
<tr>
<th>REGION</th>
<th>$p = 3.19 \times 10^{-7}$</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washington, DC</td>
<td>210.487</td>
<td>9,495</td>
<td>639.809</td>
<td></td>
</tr>
<tr>
<td>Western NC</td>
<td>42.168</td>
<td>1,528</td>
<td>583.953</td>
<td></td>
</tr>
<tr>
<td>Central NC</td>
<td>-7.242</td>
<td>5,299</td>
<td>508.778</td>
<td></td>
</tr>
<tr>
<td>Eastern NC</td>
<td>-16.532</td>
<td>658</td>
<td>556.088</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>-28.194</td>
<td>3,715</td>
<td>552.642</td>
<td></td>
</tr>
<tr>
<td>Southern NC</td>
<td>-79.971</td>
<td>1,261</td>
<td>503.350</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>-120.715</td>
<td>778</td>
<td>416.961</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENDER</th>
<th>$p = 5.20 \times 10^{-5}$</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>47.288</td>
<td>6,904</td>
<td>597.288</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-47.288</td>
<td>15,830</td>
<td>563.342</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ETHNICITY</th>
<th>$p = 0.0078$</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumbee</td>
<td>90.965</td>
<td>594</td>
<td>574.264</td>
<td></td>
</tr>
<tr>
<td>Latino</td>
<td>9.108</td>
<td>5,463</td>
<td>543.592</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>-24.240</td>
<td>3,930</td>
<td>551.246</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>-75.834</td>
<td>12,747</td>
<td>593.412</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER</th>
<th>(random effect)</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>96.509</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>349,955.1</td>
<td>13</td>
<td>569.686</td>
<td>573.651</td>
<td>22,734</td>
</tr>
</tbody>
</table>

Not selected as significant: AGE, YEAR OF BIRTH.

Age and year of birth were also submitted to the mixed model, but not found to be significant.\(^{17}\) Region is found to be the strongest predicting factor group, and it is in this category that we see the highest overall variation. Region has a log-odds range of 331.2,

\(^{17}\) In addition to age and year of birth, which were included as continuous variables, I also experimented with subdividing the speakers into age groups, as is commonly done in variationist work, but models analyzing age in this division also failed to obtain statistical significance. It has been argued that age and year of birth, which appear to be continuous variables, are actually better treated as categorical variables (as they usually are) on account of the facts that we (i.e. humans) tend to perceive and operationalize age in more categorical terms than continuous terms and that historical events are often found to cause discontinuities across ages. It is possible that different sociohistorical developments across the different locations studied here cause different salient age divisions among the communities and that better, more localized, determinations of age categories than I tested here would obtain significance.
with the mean pause duration for Washington, DC the highest, at 640 ms, while Ohio had the shortest pause duration mean of 417 ms. Notably, the model predicts the region factors in a different order than we would presume based on their mean or median values.

So, Southern NC fairly strongly disfavors longer pauses, while Central NC only barely disfavors longer pauses, despite the closeness (only a 5 ms difference) of the two groups’ means and their regional proximity.

When we examine the actual numbers for gender and ethnicity we realize that, although the model finds the factor groups significant and the log-odds ranges are fairly substantial, the actual range in values is quite small. The difference in mean pause between all females and all males for example is only 34 ms. For ethnicity, we find a 50 ms difference between the most extreme means, but we also note that the African American speakers, who have the highest mean pause, are the ones predicted by the model to favor the shortest pauses. Of the ethnic categories, the Lumbees are found to favor the longest pauses, an outcome congruent with some previous descriptions of pause in Native American languages (e.g., Philips 1976).18

At this point, it is a little difficult to make sense of overall patterns. For example, we note that, in terms of the region factor group, Washington, DC was found to favor the longest pauses, even though all the Washington, DC speakers are African American and, in terms of the ethnicity factor group, African Americans are found to favor the shortest pauses. Figure 5.4.3 illustrates the difference between the pause durations of the African American speakers in Washington, DC and the other African American speakers. A t-

---

18 I discuss the Lumbee in further depth in §7.2.1.
test indicates that the Washington, DC African Americans do, in fact, have significantly longer pauses than the African Americans from all the other regions combined ($p < 0.001$), with the mean pause for African Americans from Washington, DC at 640 ms while the other African American speakers combined have a mean of only 458 ms. In sum, it is clear that there is not a mono-dimensional effect of ethnicity in the data, and we should take from this the observation that, in general, there is likely not a direct, and singular, effect of any one social category on pause duration.

![Boxplot showing pause distribution for African Americans by region](image)

**Figure 5.4.3. Boxplot showing pause distribution for African Americans by region**

Having shown that a regression analysis finds significant social predictors for pause duration, let us return momentarily to the literature on pause – especially the
experimental literature – to note that we have established, I think, that pause duration is not only an outcome of processing activity and so forth, but that it does in fact appear to be impacted socially by such categories as regional affiliation, ethnicity, and gender, even though the effects of these social categories are far from straightforward.

5.5. The effects of pause type and filled pause adjacency on silent pause durations

So far this chapter has been concerned with pause durational differences solely in terms of the social characteristics of speakers. We now briefly consider variation in pause duration with respect to the type of pause (in terms of whether it is located at a “grammatical”, or “expected”, site or not) and the adjacency of filled pauses for a small subset of the speakers examined thus far. I have not (yet) been able to code all of the 22,734 pauses for these factors, so instead focus here on the data just from two regions, Southern NC and Ohio. Recall from the mixed-effect model above (Table 5.4.4) that the Ohio region factor had the shortest pauses (mean = 417 ms; log-odds = -121) and the Southern NC speakers had the next shortest pauses (mean = 503 ms; log-odds = -80). We examine these two groups here to better understand the effects of these linguistic factors on pause duration and also to see if the two sets of speakers are better differentiated by including these additional factors. These factors and summary statistics for the distributions of the data are described in Tables 5.5.1 and 5.5.2.

Together speakers from these two regions had 2,039 pause measurements after thresholding. In coding for the new factor groups, however, 159 pause measurements
had to be excluded for a variety of reasons (such as not being classifiable according to one or both of the factors due to unintelligible talk at the site of the pause). Thus, a total of 1,880 pauses are examined here.

Table 5.5.1. Pause duration summaries by pause type

<table>
<thead>
<tr>
<th></th>
<th>Prosodic/Grammatical Pauses</th>
<th>Non-Grammatical Pauses</th>
<th>Restart Pauses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (ms)</td>
<td>Mean (ms)</td>
<td>St.Dv. (ms)</td>
</tr>
<tr>
<td>Ohio</td>
<td>343</td>
<td>474</td>
<td>491</td>
</tr>
<tr>
<td>Southern NC</td>
<td>413</td>
<td>565</td>
<td>507</td>
</tr>
<tr>
<td>Overall</td>
<td>384</td>
<td>534</td>
<td>503</td>
</tr>
</tbody>
</table>

For the pause type factor group, as shown in Table 5.5.1, I coded the data into one of three categories, (1) “grammatical”, whether the pause came at an expected point based on the clausal or prosodic structure,19 (2) “non-grammatical”, whether it came at an unexpected point (i.e. was not predictable based on “normal” English prosody or clausal structure), or (3) “restart”, whether it was adjacent to a restart or self-repair (cf. Levelt 1989). For the most part, I based this typology on Goldman-Eisler’s (1968: 13) description of her binary schema (grammatical versus non-grammatical). However, I chose to go beyond a binary coding – of simply predicable pauses versus unpredictable pauses – and include the “restart” category, which we can consider a special case of the

---

19 The determination of whether a pause was “grammatical” or not most often was made based on the punctuation and orthography in the transcript. Only in a limited number of cases did I listen to the audio. SLAAP makes it easy to listen to the audio from the transcript and, admittedly, this should have been done to check the transcripts. However, as this section is intended here only as short exploration, I leave that for future work.
non-grammatical pause, as it seems to me a reasonable assumption that pauses that do not mark changes in the flow of talk might be cognitively different than pauses that lead to or are symptomatic of self-repair. A second difference arising from this is that I most often considered parentheticals as falling within the restart category whereas Goldman-Eisler appears to have considered them grammatical. Table 5.5.1 shows that for both sets of speakers (those from Ohio and those from Southern NC), grammatical pauses are significantly longer than non-grammatical pauses \((p < 0.001)\), and that non-grammatical pauses are longer than restart pauses, though the difference is not significant \((p > 0.05)\) in this latter case.

Table 5.5.2 displays the central tendencies of the pause duration data according to whether the pauses collocate with a filled pause or not. Following from Clark and Fox Tree’s (2002) determination that *uh* and *um* are semantically different and that *um* tends to have longer realizations and to correlate with longer pauses than *uh*, I distinguish between *uh* and *um*, making three categories for filled pause, “none”, “uh”, and “um”. As Clark and Fox Tree (2002) found, we see here that pauses collocating with *um* (mean = 797 ms) are longer than those with *uh* (mean = 648 ms), although a t-test does not find the difference significant here \((p = 0.10)\). Both types of filled pauses, however, do obtain significant differences from the “none” category of pauses not adjacent to a filled pause \((p = 0.001\) for *uh*, \(p < 0.0001\) for *um*).

---

20 There were a very few other filled pauses in the data, (er x1 and eh x6). These were among the 159 measurements that I excluded from this part of the analysis. They appeared to pattern with *uh*, but with so few tokens, it was determined that including the full set of realized filled pauses would only add unhelpful complexity to the discussion here.
Table 5.5.2. Pause duration summaries by adjacency to filled pauses

<table>
<thead>
<tr>
<th></th>
<th>No filled pause</th>
<th>Uh</th>
<th>Um</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (ms)</td>
<td>Mean (ms)</td>
<td>St.Dv. (ms)</td>
</tr>
<tr>
<td>Ohio</td>
<td>292</td>
<td>388</td>
<td>372</td>
</tr>
<tr>
<td>Southern NC</td>
<td>342</td>
<td>465</td>
<td>417</td>
</tr>
<tr>
<td>Overall</td>
<td>318</td>
<td>438</td>
<td>404</td>
</tr>
</tbody>
</table>

Figure 5.5.1 displays boxplots of the combined effects of these two factors, pause type and the presence of filled pause, on pause duration. Although the Ns are much higher for the “no filled pauses” (No FP) category, and in general filled pauses correlate with longer silent pauses, we do see here an interesting interaction where the grammatical pauses are longer than the non-grammatical and restart pauses when not collocating with a filled pause (see left-hand third of Figure 5.5.1; No FP + grammatical pause mean = 518 ms vs. No FP + non-grammatical and restart pause mean = 365 ms; \( p < 0.0001 \)), but not when there is a filled pause (\( p > 0.05 \) for both \( uh \) and \( um \)). While the differences between the grammatical categories are not significant for \( um \), we see with \( um \) (on the right-hand third of Figure 5.5.1), in fact, the opposite pattern. Here restarts (with \( um \)) are the longest pauses (mean = 1,032 ms), followed by non-grammatical pauses (mean = 781 ms), and then grammatical pauses (mean = 742 ms).

The “grammatical” pause category in Figure 5.5.1 appears fairly stable in the boxplots regardless of whether it appears with no filled pause, with \( uh \), or with \( um \), but a t-test finds that the difference between the grammatical pauses with no adjacent filled pause and the slightly longer grammatical pauses that collocate with \( um \) is just barely significant (\( p = 0.046 \)). The pattern visible in the figure with respect to restarts, that
restarts with filled pauses are longer than restarts with no filled pauses, is confirmed to be significant by a t-test ($p < 0.001$).

![Figure 5.5.1. Pause duration by pause type and adjacency to filled pause](image)

This is all interesting, but it remains to be asked how these data fit into the larger project of understanding social variation in pause realization. Figure 5.5.2 shows the same data as in Figure 5.5.1, however, here the data are divided by region. The same general patterns visible in Figure 5.5.1 appear here, although we note that the smaller Ns make the patterns less clear. The figure mostly illustrates that we do not see clear evidence of regional differences in this subset of the pause data.
To take a different approach to teasing out regional, or other social, differences in the silent pause data when pause type and filled pauses are considered, Table 5.5.3 provides the best mixed-effect model obtained through step-up/step-down analysis in Rbrul. As we would expect from the above discussion, the categories of pause type and filled pause surface as highly significant (both at $p < 0.0001$). However, the model does not contribute at all to our ability to differentiate the pausing practices of the Ohio and
Southern NC speakers. While ethnicity is found to be significant \((p < 0.01)\), further inspection shows that it is only significant due to the much longer pauses of the Lumbees.

Table 5.5.3. Mixed model for pause, with pause type and filled pause factor groups

<table>
<thead>
<tr>
<th>Filled Pause</th>
<th>(p = 1.18 \times 10^{-25})</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(U_m)</td>
<td>191.406</td>
<td>131</td>
<td>796.611</td>
<td></td>
</tr>
<tr>
<td>(U_h)</td>
<td>22.298</td>
<td>97</td>
<td>648.485</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>-213.705</td>
<td>1,652</td>
<td>438.337</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pause Type</th>
<th>(p = 2.72 \times 10^{-8})</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical</td>
<td>76.607</td>
<td>862</td>
<td>534.252</td>
<td></td>
</tr>
<tr>
<td>Restart</td>
<td>-26.789</td>
<td>203</td>
<td>441.557</td>
<td></td>
</tr>
<tr>
<td>Non-grammatical</td>
<td>-49.818</td>
<td>815</td>
<td>418.688</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>(p = 0.0046)</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumbee</td>
<td>101.373</td>
<td>565</td>
<td>567.202</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>-45.190</td>
<td>809</td>
<td>439.555</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>-56.183</td>
<td>506</td>
<td>425.540</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker (random effect)</th>
<th>Std. Dev. 89.236</th>
</tr>
</thead>
</table>

Model

<table>
<thead>
<tr>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>28,194.4</td>
<td>9</td>
<td>648.702</td>
<td>474.145</td>
<td>1,880</td>
</tr>
</tbody>
</table>

Not selected as significant: REGION, GENDER, AGE, YEAR OF BIRTH.

Removing the Lumbees from the mixed-effects model – thus leaving only African Americans and European Americans from Ohio and Southern NC – results in a model where ethnicity no longer surfaces as significant, and instead gender arises as significant (at \(p < 0.05\)).

---

\(^{21}\) Deeper examination of the data than I provide here seems to indicate that gender is obtained as significant in the Lumbee-excluded model not because there are gender differences in these data between
The data and analysis presented here do not seem to shed much light on social variation in pause practice. I had supposed (and still do in fact) that at some level beyond durational differences pausing is socially impacted. Clearly – as these data confirm – there are general patterns of pause duration based on cognitive activity (Goldman-Eisler 1968; Levelt 1989). At the same time, however, while the overall need to pause for cognitive (as well as discourse) purposes is presumably a universal of sorts, it seems to me plausible to still hypothesize that deeper examination than I have conducted here might find that different groups organize that pause time differently.

Grosjean (1980b; Grosjean and Deschamps 1975) found some evidence cross-linguistically for this. For example, comparing English and French, Grosjean explains that

the pause time ratio in the two languages is almost identical … but that this equal pause time is organized differently in the two languages: there are fewer but longer pauses in French whereas in English pauses are more numerous but shorter … Speakers of English make use of a pause slot situated inside the VP which speakers of French do not use (1980b: 307).

Of course, in the analysis presented in this section, I have not examined the locations of the pauses or the number of pauses in relation to their durations. I intended to do this and in fact have coded this subset of pauses for the syntactic location of each pause.22 This category was very complex, however, and my preliminary analyses did not point to it providing any clear picture with respect to social variation. I decided that describing its

---

22 Grosjean and Collins (1979; see also Grosjean 1980a) examined the influence of constituent and syntactic structures on pause frequency and duration and indicate that the location of a pause is an important linguistic factor on its duration (and its occurrence in the first place). I had hoped to examine social variation in these terms as well, but must save that project for future work.
analysis would be too space consuming here and would still fail to speak to my overall goals or Grosjean’s findings. Possibly, the analysis of this section would have obtained more striking findings had I examined a larger subset of the data or had I not focused on the two most similar regional groups, in terms of the earlier findings. I end this section, then, acknowledging this and hoping to have highlighted more future directions for this work than I have contributed to a clearer picture of the variation found in the pause data here.

5.6. In closing: Variation in pause

I ended §5.2 by asking whether we could determine whether social variation in pause realization was meaningful. §5.4 established, I believe, that social variation in pause realization exists – and, in fact, may be extensive, as indicated by the fact that the region-factors of Central NC and Southern NC were found to predict different pause durations despite their geographical proximity and their very similar mean pause values. However, this finding does not answer the question of whether the variation in pause realization is necessarily meaningful in the sense that it impacts or is meaningful to interlocutors as they conduct their daily business over the courses of their linguistic lives.

One way of answering this question has nothing to do with the quantitative work undertaken here. Reconsidering some of the work from areas of discourse analysis and language and law, we could readily argue that – yes – variation in pause is extremely meaningful. Even just within the language and law literature, for example, we see
evidence of this: Norma Mendoza-Denton (1995) demonstrated the importance of gap-length in understanding the Clarence Thomas-Anita Hill hearings; my own work on pause in deposition (Kendall 2007b) has illustrated, I hope, the potential for misunderstanding if pauses are not recorded in court reported transcripts; and, as a final example, Anne Graffam Walker (1985) demonstrated the extent that hesitancy (i.e. pausing) plays on lawyers’ impressions and memory of witnesses’ testimony.

But, do the findings of this quantitative pause analysis bear on the question? Quené (2007) evaluated the question of what relative change is necessary before hearers perceive a difference in timing. This minimum perceivable change – called the just noticeable difference (or, JND) – is a useful way to think about meaningfulness of pause variation. So, recall that gender was found by the mixed-effect model in Table 5.4.4 to be a significant factor in pause duration, but there was only a 34 ms difference between the mean pause durations for the males and females. Is this meaningful? One way to answer this would be to say: Only if it is perceivable.23 Quené mostly focused on changes in rate of speech – and I do not know of work that has tested the JND specifically for pause – but his findings can be brought to bear on the present work. Quené (2007) found a

\[ \text{JND of about 5\% of the base tempo of a speech utterance.} \]
\[ \text{Tempo variations exceeding this DL \cite{Quené2007} are likely to be noticeable, and relevant in speech communication (2007: 360).} \]

23 Of course, I acknowledge that much sociolinguistic work, such as the research on near-mergers (e.g., Labov 1994), has demonstrated that perceivability is not the only requirement for a difference to be linguistically important.
Assuming (and with the acknowledgement that this is an area needing further experimental testing) that a JND of about 5% exists for pause durational differences, then the difference between the mean durations of males and females of 34 ms – which is 5.7% – is likely perceivable, and following Quené’s logic, “relevant in speech communication” (2007: 360). Meanwhile, while readers may find it a leap to use Quené’s findings on the JND for rate of speech to bear on these pause data, note that much of the work reviewed by Rochester (1973) and discussed by Crown and Feldstein (1985) indicates that hearers are quite perceptive of pause differences when asked about rate of speech differences. I introduced Quené’s work here to quantitatively nuance the general point – variation in pause appears quite meaningful.

We now turn to a similar analysis of speech rate, before considering both features – pause and speech rate – together in more nuanced ways.
6. Toward a quantitative sociolinguistic analysis of speech rate

6.1. Introduction

While speech rate, as an area of linguistic inquiry, has not been problematized in the same way as pause (again cf. Macaulay 2002), analyses of speech rate have remained fairly marginal pursuits, especially within sociolinguistic domains. Nonetheless, a few studies have pursued speech rate as a sociolinguistic phenomenon, and have yielded provocative findings that indicate that speech rate may be a rich area for variationist study. Before turning to a brief review of work that has been conducted along these lines, I first provide some terminological clarifications.

6.2. Defining speech rate terminology

What I have been terming speech rate is often referred to as speaking rate or articulation rate in the literature. In these more precise terms, speaking rate is often used to refer to a measure that includes pauses, while articulation rate refers to a measure with pauses longer than a certain threshold, say 150 – 250 ms, omitted (cf. Robb, Maclagan, and Chen 2004). The measure used in the present study (as described in §6.4) is equivalent to articulation rate as it is computed from speech uninterrupted by pauses, where a pause is defined as any silence longer than ~ 60 ms.
For the course of this study, I will use the term *speech rate* to refer to *articulation rate*, feeling that, for our purposes, the term is clearer and more obvious to readers. In the following section, where I discuss previous findings on speech rate, I will use the term *speaking rate* for the measure that includes pauses, but will continue to use the term *speech rate* as synonymous with *articulation rate*. I will occasionally use *rate of speech* as a more general term, to avoid indication of a particular quantitative metric.

To reiterate, for the data presented here, speech rate (again, = articulation rate) excludes all pauses over 60 ms in duration.

### 6.3. Research on speech rate

For many of the primary psycholinguistic researchers of the temporal sequencing of speech (such as Goldman-Eisler, Deese, O’Connell, and Kowal) speech rate has been studied as a secondary phenomenon after pause. This is likely a result of these researchers having a primary interest in the window that speech timing features (such as pause and speech rate) might lend into language planning and production and the finding early on (cf. Goldman-Eisler 1968) that speech rate (measured via articulation rate) exhibited little change based on such factors as the difficulty of the speaking task. In fact, Goldman-Eisler’s (1954, 1961, 1968) principal experimental finding about speech rate was that variation in a speaker’s articulation rate is mainly influenced by practice and repetition – with practiced talk spoken significantly faster than spontaneous talk. She
writes that articulation rate “thus becomes an efficient and unequivocal indicator of habit strength only” (Goldman-Eisler 1968: 26).

Goldman-Eisler (1954, 1961) further found that

the speed of the actual articulation movements producing speech sounds occupies a very small range of variation (4.4 to 5.9 syllables per second were obtained from speech uttered during interviews) while the range of pause time in relation to speech time was five times that of the rate of articulation (Goldman-Eisler 1961: 171).

In other words, according to Goldman-Eisler (e.g., 1968: 26), what hearers perceive as changes in the rate of speech is primarily the result of changes in pausing by the speaker.\(^1\)

While this stance has been supported by others (e.g., Grosjean 1980b), the notion that it is primary has also been refuted. Miller, Grosjean, and Lomato (1984) demonstrated that variation in speech rate is significant on its own, even within single speech events. In fact, Miller et al. (1984) argue further that speech rate variation was significantly underappreciated in the earlier work of scholars like Goldman-Eisler (and Grosjean 1980b); the high degree of intra-speaker variability in speech rate obscures inter-speaker variation when aggregate data are used to represent speakers’ speech rate. That is, “articulation rate typically has been measured over large stretches of speech, such that the local variation characterized by the peaks and troughs that can be seen … is neutralized” (Miller et al. 1984: 222).

So, despite Goldman-Eisler’s finding of mostly invariance at the speaker-level, later scholars have focused on intra-speaker speech rate variation. Deese (1984), for

\(^{1}\) In fact, it is this finding – the importance of pause on the perception of speech rate – that motives the general distinction between speaking rate – the measure including pause – and articulation rate (or what I’m terming just speech rate) – the measure excluding pause. See §6.2, above, for a discussion of this terminology.
example, reported a “normal” speaking rate for conversational speech to be between 5 and 6 syllables per second, but further argued that speakers tend to speed up toward the end of utterances as strategies to keep the floor. At the same time, other researchers have found the opposite, that the last few words of an utterance are the longest. For example, Yuan, Liberman, and Cieri (2006) show that speakers slow down at the end of utterances, a result that conflicts with (or perhaps nuances) Deese’s (1984) claim.

The psycholinguistic tradition, discussed thus far, has focused on intra-speaker variation in speech rate. From a sociolinguistic perspective, we are of course interested in depth in inter-speaker variation, and, interestingly here, there has been a lot of disagreement when it comes to the significance of inter-speaker variation in speech rate. Goldman-Eisler (cf. 1968) found that her subjects showed a great deal of individual differences in their overall speech rates. Deese, on the other hand, declared rather boldly, “few native-born speakers of the standard dialect of English vary much in their rate of speaking” (Deese 1984: 105). I will return below to considering possible explanations for this disagreement.

Despite the relative lack of interest in pursuing social variation in speech rate by the foundational psycholinguists, it appears that speech rate has been examined by a wider range of research groups than pause due, at least in part, to its relevance for addressing speech disorders. Researchers have addressed normative speech rates for specific language varieties (e.g., Block and Killen 1996 on Australian English; Robb, Maclagan, and Chen 2004 on New Zealand English and American English), issues with respect to specific populations (e.g., Van Borsel and De Maesschalck 2008 on
transsexuals’ speech), and on specific articulatory and production hypotheses (e.g., Tsao and Weisner 1997). I will not address this entire broad literature here. Instead, I briefly discuss some relevant findings from a select few papers.

Speech rate differences between speakers above the level of the individual have been examined to some extent, though primarily, it seems, in terms of regional differences. At a macro-regional level, Robb, Maclagan, and Chen (2003), for example, compared speech rates between 40 speakers of New Zealand English and 40 speakers of American English and demonstrated that “not all varieties of English are spoken at the same rate” (Robb et al. 2003: 12). Regional differences in speech rate have also been found within American English by some researchers (e.g., Salmons, Jacewicz, and Fox 2008), but not by others (e.g., Ray and Zahn 1990, although these authors note that that null result is surprising to them). Yuan et al. (2006) confirm the impression that there have been conflicting findings regarding variation in speech rate at the regional dialect level.

The same can be said of gender-based variation in speech rate – much research points to males speaking faster than females, but it is often weak or mitigated evidence. Yuan et al. (2006) note this finding, but also that the difference between males and females – albeit statistically significant – is very minor. Salmons et al. (2008) point out and confirm that general findings indicate males often speak faster than females. However, Salmons et al. (2008), in their comparison of speech rate in Wisconsin and Western North Carolina find that gender interacts with regional dialect; they find no effect of gender on speech rate in Wisconsin, or for spontaneous speech in the North
Carolina mountains, but in their data the read speech by women is significantly slower in Western North Carolina than the read speech by males. Clopper and Smiljanic (2007), examining differences between Midland American English and Southern American English, find no significant differences for gender or dialect in speech rate. Yuan et al. (2006), and others (e.g., Quené 2008), find some evidence that older speakers speak more slowly than young speakers.

In sum, studies of speech rate have found significant differences at the individual level (e.g., Goldman-Eisler 1968) and between macro-level varieties (e.g., Robb et al. 2003). In terms of finer-level sociolinguistically-relevant differences, however, findings have been fairly contradictory, with some researchers finding significant differences at the regional and gender levels (e.g., Salmons et al. 2008) and others finding no significant differences (e.g., Clopper and Smiljanic 2007).

One reason for the contradictory findings – beyond Goldman-Eisler’s (e.g., 1968) understanding of speech rate as highly idiosyncratic – may be related to a strong correlation between utterance length (in terms of numbers of syllables or words per utterance) and speech rate. Quené (2008) investigated the effect of “anticipatory shortening” – the tendency of utterances with more syllables to be spoken with shorter syllables – on a larger investigation of regional, gender, and age differences on speech rate in Dutch dialects. He found that, indeed, utterance length has a highly significant effect on speech rate and that by including that within-speaker factor (in a mixed-effect model analysis) the between-speaker factors of age and gender become mitigated. I will return to this below (in §6.5.1).
A second reason for conflicting results in the previous literature may relate more simply to the varied measures (see §6.4) used for speech rate. That is, it seems possible that simple mathematical problems of precision of measurement (such as orders of magnitude errors, and rounding errors) hide for some studies what might otherwise be found to be significant variation. I began this section by quoting Goldman-Eisler’s (1961: 171) report that “a very small range of variation (4.4 to 5.9 syllables per second”) was found for articulation rate. But, what is centrally at issue here is that we might disagree with the categorization of a 1.5 σ/sec range as “a small range of variation.” While perception questions are outside the scope of this dissertation, when considering speech features like speech rate (and pause), we must revisit the discussion at the end of the previous chapter (§5.6) and ask to what degree differences in these features are perceptible to listeners.\(^2\) As a reminder from §5.6, Quené (2007) found that hearers perceived rate of speech changes greater than about 5% (i.e. that the *just noticeable difference*, or JND, is \(\approx 5\%\)). This indicates that hearers may perceive differences in speech rate on the order of \(\pm 0.25 \sigma/\text{sec}\) (based on an average speech rate of somewhere around \(5 \sigma/\text{sec}\)). In other words, differences in speech rates between 4.4 \(\sigma/\text{sec}\) and 5.9 \(\sigma/\text{sec}\) would be quite noticeable and should not be considered “a small range of variation” (Goldman-Eisler 1961: 171) at all.\(^3\)

---

\(^2\) That is, to what extent does statistically significant variation matter if it is below the level of speakers’ perception (Labov 1994; Vaughn 2008)?

\(^3\) An interesting question that relates to the work of this chapter, but which I must leave for future work, is whether there is variation in JND relating to hearers’ native dialects, or whether JND is a purely physiological trait.
6.4. Automating speech rate measurements

As we have seen, speech rate can be measured and discussed in a variety of units. For example, *words per minute* (wpm) is a common metric, as is *syllables per minute* (spm or σ/m). Yuan et al. (2006) even discuss characters per minute when they discuss speech rate in Chinese. For the work presented here, I report all speech rate measures in terms of *syllables per second* (σ/sec). The syllables per second measure seems to me to give a more precise measure than words per minute, and to indicate the higher degree of accuracy. A number of other scholars (e.g., Clopper and Smiljanic 2007, Salmons et al. 2008, and often Miller et al. 1984) also use this unit. Some scholars (e.g., Tsao and Weismer 1997) discuss speech rates in terms of *milliseconds per syllable* (ms/syll) – in fact, Miller et al. (1984) switch between ms/syll and σ/sec – but again, I prefer σ/sec and will be consistent in its use.

Similar to the implementation of a pause analysis tool, I have developed a feature in SLAAP that automatically calculates the speech rate for each line in a transcript. For the determination of speech rate, the software selects all transcript lines for utterances that have durations between a low threshold of 0.5 seconds and a high threshold of 5 seconds and that do not match certain *don’t count* criteria. No transcript lines containing filled pauses (*uh* or *um*) are used. Transcript lines containing SLAAP transcript convention-characters for unsure transcription (/…/), speaker overlap ([…]), or non- or semi-linguistic noises (<…>; e.g., <laughter>, <clap>, <cough>) are also not used for speech rate analysis. Lines containing speech phenomena like false starts or restarts (typically indicated with dashes, -) are included in the analysis, and the syllable counting
algorithm attempts to count those based on their orthography. For example, the utterance transcribed as “did- did- did a lot of- lot of traveling around, right” (phb_HH1_600_1200: 364) is counted as 14 syllables and the utterance transcribed as “it run, I c-” (bee0010a_0_2756: 335) is counted as 4 syllables. See §3.3 and the SLAAP User Guide (2008) for more information about SLAAP’s transcription conventions.

An algorithm then determines speech rate by counting syllables in the orthographic transcript and dividing that count by the duration of the transcript line. SLAAP then reports a syllable per second measure for each selected line, and computes a median value over all selected lines. As reported in Kendall (2007a: footnote 9), the automatic syllable counting algorithm has been tested to be about 77% accurate. However, less than 2% of the wrong cases are off by more than one syllable, so these wrong cases are rarely considered problematic.

In fact, for the determination of speech rate, SLAAP allows us the manual correction or removal of incorrect syllable counts, but since we are examining large segments of talk here, speech rate (and pause duration) medians have been calculated from very high _Ns and are quite stable. As an example, the median speech rate determination for Alayna, one of the African American females from Washington, DC obtained 5.25 syllables per second based on 1093 lines of speech; halving the number of lines used (i.e. taking the median of the first 547 measurements), however, we still obtain a value of 5.26 σ/sec (a difference of 0.01 σ/sec; _p > 0.05). Randomly selecting 546 measurements from Alayna’s 1093 total measurements, we obtain a median value of 5.20 σ/sec (a difference of 0.05 σ/sec; _p > 0.05). Since it seemed possible that this could be
more problematic for speakers for which we have fewer measurements, I tested each speaker’s total speech rate measurements against a random sampling of half of their measurements. In doing so, I obtained no significant differences for any speaker.\footnote{While no differences were found to be significant in these randomizing comparisons, some speakers’ values did change more than others. Nonetheless, t-tests obtained \( p \) values between 0.995 and 0.118 (median = 0.661), all clearly well above the threshold for significance of 0.05.}
The SLAAP speech rate analysis tool is illustrated in Figure 6.4.1. I then, as for the pause data, developed a script in R, which, using the speech rate analysis tool in SLAAP, extracts the data and allows for the large-scale processing and comparing of all 104 speakers.\(^5\)

6.5. Analysis and discussion

We turn now to the analysis of the overall speech rate data from the 104 speakers introduced in §4.2. Figure 6.5.1 shows the distribution of all 23,871 speech rate measurements. With the exception of a small peak at around 1.5 \(\sigma/\text{sec}\), speech rate distributes normally. The overall mean over all 23,871 measurements is 4.59 \(\sigma/\text{sec}\) with a standard deviation of 1.56 \(\sigma/\text{sec}\).

\(^5\) While I believe that the syllable counting errors were marginally important, most of these errors appeared to result in syllable per second measurements that were unrealistically high. For this reason, the R script was set to ignore all measurements above 10 \(\sigma/\text{sec}\) as an additional safeguard in the belief that normal human speech will never be that fast. It is believed that this modification removed a proportion of potential error as well.
Figure 6.5.1. Distribution of speech rate measurements

Figure 6.5.2 shows boxplots for all speakers, ordered by median speech rate. The same colors used earlier in Figure 5.4.2 are again used here to indicate speaker ethnicity and gender: Dark green is used to display white males, light green for white females, dark blue for black males, light blue for black females, light brown for Latino males, yellow for Latina females, red for Lumbee males, and pink for Lumbee females. The mean speech rate across all speakers (calculated from the speakers’ individual medians) is 4.79 \( \sigma/\text{sec} \).

While it was not possible to “eyeball” patterns in the pause data of the last chapter, readers with a color version of Figure 6.5.2 will immediate note that there appear to be fairly clear patterns in the distribution according to ethnicity and gender. Blue (especially light blue) and yellow bars are heavily distributed on the left-hand side of the plot.
Figure 6.5.2. Speech rate by speaker, ordered by median speech rate
Green bars, both light and dark, appear to cluster more on the right-hand side. In other words, African American females (light blue) and Latinos of both genders (yellow and light brown) appear in Figure 6.5.2 to have lower speech rates in general than white speakers (light and dark green).

Tables 6.5.1 – 6.5.3 present the overall speech rate measurement data by ethnicity, gender, and region, ordered by median speech rate. As before, the complete data are available in the Appendix.

### Table 6.5.1. Speech rate data by ethnicity

<table>
<thead>
<tr>
<th></th>
<th>By Measurement</th>
<th>By Speaker (based on speaker’s median speech rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of tokens</td>
<td>Median</td>
</tr>
<tr>
<td>European American</td>
<td>3,618</td>
<td>5.25</td>
</tr>
<tr>
<td>Lumbee</td>
<td>742</td>
<td>4.78</td>
</tr>
<tr>
<td>African American</td>
<td>14,922</td>
<td>4.49</td>
</tr>
<tr>
<td>Latino</td>
<td>4,589</td>
<td>4.43</td>
</tr>
<tr>
<td>Overall</td>
<td>23,871</td>
<td>4.61</td>
</tr>
</tbody>
</table>

### Table 6.5.2. Speech rate data by gender

<table>
<thead>
<tr>
<th></th>
<th>By Measurement</th>
<th>By Speaker (based on speaker’s median speech rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of tokens</td>
<td>Median</td>
</tr>
<tr>
<td>Male</td>
<td>6,741</td>
<td>5.03</td>
</tr>
<tr>
<td>Female</td>
<td>17,130</td>
<td>4.46</td>
</tr>
<tr>
<td>Overall</td>
<td>23,871</td>
<td>4.61</td>
</tr>
</tbody>
</table>

142
Table 6.5.3. Speech rate data by region

<table>
<thead>
<tr>
<th>Region</th>
<th># of tokens</th>
<th>Median</th>
<th>Mean</th>
<th>St. Dev.</th>
<th># of spks</th>
<th>Mean tokens per spkr</th>
<th>Median</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Texas</td>
<td>2,916</td>
<td>5.18</td>
<td>5.13</td>
<td>1.47</td>
<td>24</td>
<td>122</td>
<td>5.08</td>
<td>5.13</td>
<td>0.53</td>
</tr>
<tr>
<td>Ohio</td>
<td>679</td>
<td>5.15</td>
<td>5.16</td>
<td>1.49</td>
<td>7</td>
<td>97</td>
<td>5.12</td>
<td>5.23</td>
<td>0.51</td>
</tr>
<tr>
<td>Southern NC</td>
<td>1,530</td>
<td>5.03</td>
<td>5.01</td>
<td>1.53</td>
<td>19</td>
<td>81</td>
<td>4.76</td>
<td>5.04</td>
<td>0.63</td>
</tr>
<tr>
<td>Eastern NC</td>
<td>686</td>
<td>4.98</td>
<td>4.95</td>
<td>1.70</td>
<td>5</td>
<td>137</td>
<td>5.11</td>
<td>5.02</td>
<td>0.98</td>
</tr>
<tr>
<td>Western NC</td>
<td>1,861</td>
<td>4.53</td>
<td>4.53</td>
<td>1.51</td>
<td>8</td>
<td>233</td>
<td>4.20</td>
<td>4.22</td>
<td>0.65</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>11,025</td>
<td>4.51</td>
<td>4.45</td>
<td>1.52</td>
<td>12</td>
<td>919</td>
<td>4.37</td>
<td>4.43</td>
<td>0.57</td>
</tr>
<tr>
<td>Central NC</td>
<td>5,174</td>
<td>4.29</td>
<td>4.33</td>
<td>1.56</td>
<td>29</td>
<td>178</td>
<td>4.58</td>
<td>4.52</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>23,871</td>
<td>4.61</td>
<td>4.59</td>
<td>1.56</td>
<td>104</td>
<td>230</td>
<td>4.76</td>
<td>4.79</td>
<td>0.72</td>
</tr>
</tbody>
</table>

These data were analyzed using a similar mixed-effect model in Rbrul as was used in Chapter 5.6 The results of the best model from a step-down analysis are shown in Table 6.5.4.7 Here the model selects region (p < 0.001) and ethnicity (p < 0.01) as significant. Speakers from Texas, Southern NC, Ohio, and Eastern NC favor faster speech rates, whereas speakers from Central NC, Western NC, and Washington, DC favor slower speech rates.

Meanwhile, European American speakers have the highest (predicted) speech rate, followed by African Americans, and Latino speakers have to slowest (predicted) speech rate. The model also selects age as significant (p < 0.05), but we note the log-odds influence of age is rather small, with increases in age predicting minor decreases in

---

6 See §5.4 for a brief discussion of mixed-effects modeling and the rationale for its use here.
7 The log-odds values from models such as Table 6.5.4 must be considered relative to the local model. The log-odds values in Table 5.4.4 were larger than they are here because they were with respect to the pause durational values, which, in milliseconds, have means in the 500s compared to the single digit values for speech rate.
speech rate. Year of birth and gender were also submitted to the mixed model analysis, but not found to be significant.\(^8\)

### Table 6.5.4. Mixed-effect model for all speech rate data

<table>
<thead>
<tr>
<th>REGION</th>
<th>( p = 0.000381 )</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texas</td>
<td>0.565</td>
<td>2916</td>
<td>5.129</td>
<td></td>
</tr>
<tr>
<td>Southern NC</td>
<td>0.343</td>
<td>1530</td>
<td>5.012</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>0.201</td>
<td>679</td>
<td>5.165</td>
<td></td>
</tr>
<tr>
<td>Eastern NC</td>
<td>0.170</td>
<td>686</td>
<td>5.046</td>
<td></td>
</tr>
<tr>
<td>Central NC</td>
<td>-0.238</td>
<td>5174</td>
<td>4.329</td>
<td></td>
</tr>
<tr>
<td>Western NC</td>
<td>-0.490</td>
<td>1861</td>
<td>4.534</td>
<td></td>
</tr>
<tr>
<td>Washington, DC</td>
<td>-0.552</td>
<td>11025</td>
<td>4.452</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ETHNICITY</th>
<th>( p = 0.00297 )</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0.348</td>
<td>3618</td>
<td>5.222</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.129</td>
<td>14922</td>
<td>4.465</td>
<td></td>
</tr>
<tr>
<td>Lumbee</td>
<td>-0.183</td>
<td>742</td>
<td>4.803</td>
<td></td>
</tr>
<tr>
<td>Latino</td>
<td>-0.294</td>
<td>4589</td>
<td>4.455</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AGE</th>
<th>( p = 0.0353 )</th>
<th>(continuous predictor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>-0.006</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.598</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84,932.88</td>
<td>13</td>
<td>4.885</td>
<td>4.588</td>
<td>23,871</td>
</tr>
</tbody>
</table>

Not selected as significant: YEAR OF BIRTH, GENDER.

Recall that Table 6.5.2 indicated that there is about a 0.5 \( \sigma/\text{sec} \) difference between males’ and females’ overall speech rates. Further, in Figure 6.5.1, above, we observed

\(^8\) The model presented here is the best outcome of a step-down mixed model analysis. A step-up analysis finds that year of birth (and not age) is significant (at the \( p < 0.05 \) level) in addition to region and ethnicity. However, in this step-up best model, ethnicity is only significant at the \( p < 0.05 \) level and the strength of the year of birth effect is the same as that of age in the step-down model. The decrease in significance favored the selection of the step-down model, which is discussed in the remainder of this analysis.
some striking indications that gender – and in particular its interaction with ethnicity – may be an important factor for speech rate despite the fact that gender does not surface as significant in the mixed-effect model in Table 6.5.4. To examine this more closely, Figure 6.5.3 displays the median speech rates for individual speakers organized by ethnicity and gender. 9

Figure 6.5.3 strengthens the impression that there appear to be some differences by gender depending on ethnicity despite the model’s outcome. European males and females pattern similarly, but African American, Latina, and Lumbee females have speech rates centered around 4.5 $\sigma$/sec while all other groups have speech rates closer to 5 $\sigma$/sec. In fact, if we group European American females with males from all ethnicities and compare this group to the remaining females we obtain a mean speech rate of 5.01 $\sigma$/sec for the males + European American females and 4.51 $\sigma$/sec for the non-European American females, with a t-test obtaining significance ($p < 0.001$). Disregarding European Americans altogether, we obtain mean speech rates of 4.51 $\sigma$/sec 10 for the females and 4.85 $\sigma$/sec for the males – though the difference is no longer found to be significant by a t-test ($p = 0.065$).

---

9 The plotting shapes used earlier in Figure 4.2.1 are re-used here to help indicate the distribution by region in this figure. The dashed line in each plot indicates the mean speech rate (based on the individual medians) for that gender + ethnicity category.

10 This value is the same as above, since this group of speakers didn’t include European Americans in the first place.
Figure 6.5.3. Speech rates by ethnicity and gender
In short, it appears that the mixed-effects model of Table 6.5.4, which did not account for interactions, failed to capture the important role of gender in its interaction with ethnicity. The role of gender itself is complexly mitigated by the European Americans, where both males and females appear to have similarly patterning speech rates, and by the fact that without the European Americans, we do not obtain significance in comparing the remaining speakers by gender.

Table 6.5.5. Mixed-effect model for speech rate, examining the gender + ethnicity interaction

<table>
<thead>
<tr>
<th>REGION</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern NC</td>
<td>0.352</td>
<td>1530</td>
<td>5.012</td>
</tr>
<tr>
<td>Texas</td>
<td>0.340</td>
<td>2916</td>
<td>5.129</td>
</tr>
<tr>
<td>Ohio</td>
<td>0.253</td>
<td>679</td>
<td>5.165</td>
</tr>
<tr>
<td>Eastern NC</td>
<td>0.003</td>
<td>686</td>
<td>5.046</td>
</tr>
<tr>
<td>Central NC</td>
<td>-0.245</td>
<td>5174</td>
<td>4.329</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>-0.250</td>
<td>11025</td>
<td>4.452</td>
</tr>
<tr>
<td>Western NC</td>
<td>-0.452</td>
<td>1861</td>
<td>4.534</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENDER + ETHNICITY</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male + White</td>
<td>0.387</td>
<td>2633</td>
<td>5.300</td>
</tr>
<tr>
<td>Male + Black</td>
<td>0.244</td>
<td>2131</td>
<td>4.641</td>
</tr>
<tr>
<td>Female + White</td>
<td>0.237</td>
<td>985</td>
<td>5.011</td>
</tr>
<tr>
<td>Male + Latino</td>
<td>0.010</td>
<td>1540</td>
<td>4.909</td>
</tr>
<tr>
<td>Female + Black</td>
<td>-0.096</td>
<td>12791</td>
<td>4.435</td>
</tr>
<tr>
<td>Male + Lumbee</td>
<td>-0.115</td>
<td>437</td>
<td>4.927</td>
</tr>
<tr>
<td>Female + Latina</td>
<td>-0.242</td>
<td>3049</td>
<td>4.226</td>
</tr>
<tr>
<td>Female + Lumbee</td>
<td>-0.425</td>
<td>305</td>
<td>4.626</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKER</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(random effect)</td>
<td>0.607</td>
</tr>
</tbody>
</table>

Model

<table>
<thead>
<tr>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>84,933.25</td>
<td>16</td>
<td>4.667</td>
<td>4.588</td>
<td>23,871</td>
</tr>
</tbody>
</table>

Not selected as significant: AGE, YEAR OF BIRTH.
Investigating the effect of the gender + ethnicity interaction in a mixed-effect model, we obtain the model presented in Table 6.5.5. Looking at the gender + ethnicity factor group, we see here that it is not only the European American females who cluster out of their gender grouping by favoring faster speech rates, but also the Lumbee males, who pattern more like the females overall, in favoring slower speech rates according to the model (despite their fairly high mean speech rate). Note that the gender + ethnicity category, however, is only barely significant ($p = 0.048$), a fact that helps explain why gender did not appear significant in the earlier model (in Table 6.5.4). Attempting to come to terms with the role of gender as a predictor of speech rate, it appears that males overall might favor faster speech rates than females, but that certain ethnicities diverge from this overall pattern, and do so to a great enough extent to limit the statistical significance of that more general claim.

Considering the fairly mixed findings and claims in the existing literature, the results of these analyses are not surprising. In many ways, the strong significance of region in the mixed-effects models is complicated by the fact that there is no easily interpretable pattern to the ordering of the region factors. In fact, note that some of the region factors fluctuate in their positions with respect to one another in the two models presented above (Tables 6.5.4 and 6.5.5). Both models do agree in their placement, however, of the Ohio speakers examined here between the faster speech rate speakers of

---

11 The model in Table 6.5.5 is the result of the step-down analysis. The step-up analysis again obtains a slightly different model, here with age being realized as barely significant (but also with only a very slight effect). The step-up analysis has higher df and otherwise is no better than the step-down model so I have only presented the step-down model here. Anyway, we are primarily interested in the gender + ethnicity interaction, which is sufficiently illustrated by either model.
Southern NC and the slower speech rate speakers of Eastern, Central, and Western NC, and we can make the substantive observation that this ordering is contrary to many stereotypes about fast talking Northerners and slow talking Southerners (see, for example, Clopper and Smiljanic 2007). At the same time, recalling from the model of Table 5.4.4 in the previous chapter that the Ohioan speakers obtained the shortest pauses, this also perhaps supports Goldman-Eisler’s claim (cf. 1968) that perceived speech rate differences may be in actuality largely the effect of differences in pause frequency and duration.

Meanwhile, from a different perspective, we may interpret the fact that speakers from Southern NC and speakers from Central NC surface so far apart in the model’s predictions as a good indication that there can be substantial differences in speech rate even in relatively proximate locations. In fact, the mean speech rate for Central NC of 4.3 $\sigma$/sec is 13.6% below that of the 5.0 $\sigma$/sec rate of Southern NC. This sizeable difference is well above the JND of 5% (Quené 2007) and should indicate that there are noticeable differences in the speech rate of speakers (at least at the aggregate level) between these two areas.

6.5.1. *The effect of utterance length on speech rate*

Having established that speech rate varies according to a number of non-linguistic factors, we will now return to the claim that utterance length – in terms of numbers of words or syllables – may in fact be of central importance in understanding and modeling
speech rate (Yuan et al. 2006; Quené 2008). Figure 6.5.4 displays the aggregate data by utterance length and shows the clear tendency that speech rate measures, on average, increase with the length of the utterance. The scatter among the higher utterance lengths, in Figure 6.5.4, can likely be attributed to the low Ns for utterances longer than about 20 syllables. This is illustrated in Figure 6.5.5, which indicates the number of measurements for utterances of each length.

![Mean Speech Rate by Utterance Length](image)

**Figure 6.5.4. Speech rate as a function of utterance length**

It seems clear from these figures that, at least in the aggregate, speech rate is to a large degree a function of utterance length. Meanwhile, we also see that the majority of utterances have between one and ten syllables, with a mean utterance length of 6.96 syllables and a median of 6 syllables.
To examine the effect of utterance length on these data, I revisit the model presented above in Table 6.5.5, but here including the continuous factor of utterance length. The results of this new model are shown in Table 6.5.6. Interestingly, region, which was the most significant factor group in the previous models, is no longer found to be significant. Instead, utterance length is obtained as extremely significant, with a 1-syllable change in utterance length corresponding to a +0.173 change in the log-odds. In addition to its significance, this effect is quite strong, with a change of 2-3 syllables in length causing a greater predictive effect than any gender + ethnicity group.

\[^{12}\text{I use the model including the gender + ethnicity interaction because it appears to better model the observable patterns in the data.}\]
### Table 6.5.6. Mixed-effect model for speech rate including utterance length

<table>
<thead>
<tr>
<th>Utterance Length</th>
<th>p ≈ 0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds</td>
</tr>
<tr>
<td>+1</td>
<td>0.173</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender + Ethnicity</th>
<th>p = 0.00686</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds</td>
</tr>
<tr>
<td>Male + White</td>
<td>0.303</td>
</tr>
<tr>
<td>Female + White</td>
<td>0.224</td>
</tr>
<tr>
<td>Male + Black</td>
<td>0.193</td>
</tr>
<tr>
<td>Male + Latino</td>
<td>0.146</td>
</tr>
<tr>
<td>Male + Lumbee</td>
<td>0.044</td>
</tr>
<tr>
<td>Female + Latina</td>
<td>-0.085</td>
</tr>
<tr>
<td>Female + Black</td>
<td>-0.335</td>
</tr>
<tr>
<td>Female + Lumbee</td>
<td>-0.490</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker</th>
<th>(random effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std. Dev.</td>
</tr>
<tr>
<td></td>
<td>0.560</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
</tr>
<tr>
<td>75,016.34</td>
</tr>
</tbody>
</table>

Not selected as significant: REGION, AGE, YEAR OF BIRTH.

We may also note that the ordering of the gender + ethnicity categories is slightly different than in Table 6.5.5’s model, but that for all intents and purposes the main findings regarding gender and ethnicity above are not only confirmed but found to be more significant. Here other than European American females, all females favor slower speech rates and all males favor faster speech rates. Moreover, for both genders Lumbees have the slowest (predicted) speech rates.$^{14}$

---

$^{13}$ The total N is two higher here than in previous models because in late stages of analysis I edited the transcript for a Lumbee speaker and changed two utterances from unsure transcription to sure transcription. These two utterances then became available for analysis and were selected by the automated procedure when I re-downloaded the data from SLAAP in order to capture the utterance length information. Quick test have shown that these two additional measurements do not impact the earlier models.

$^{14}$ The fact that including utterance length as a predictor of speech rate obtains a model where the region factor group – previously the strongest predictor of speech rate – is no longer found to be significant is worth greater consideration. Where does the influence of region go in the model of Table 6.5.6? Models
6.6. In closing: Variation in speech rate

To summarize the findings of this chapter, speech rate – a measure of the syllables per second of a speaker’s actual phonation – appears to have some social correlates (most interestingly an interaction of gender and ethnicity), but is principally that consider utterance length (again, in terms of syllables per utterance) as the dependent variable, in fact, obtain region and gender + ethnicity as highly significant predictors (shown below). In addition to being significant, we see that the effects are quite strong, with ranges of about 2.5 log-odds for both region and gender + ethnicity. There is also close to a 2-syllable range in the mean utterance lengths by region and almost a 4-syllable range in the mean lengths by gender + ethnicity. We also note that the order of factors for region is roughly the same as found for the speech rate model without utterance length (Table 6.5.5), but that the gender + ethnicity order is substantially different.

Mixed-effect model for influences on utterance length

<table>
<thead>
<tr>
<th>REGION</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texas</td>
<td>0.999</td>
<td>2916</td>
<td>7.569</td>
</tr>
<tr>
<td>Ohio</td>
<td>0.854</td>
<td>679</td>
<td>8.297</td>
</tr>
<tr>
<td>Southern NC</td>
<td>0.794</td>
<td>1532</td>
<td>8.343</td>
</tr>
<tr>
<td>Eastern NC</td>
<td>0.525</td>
<td>686</td>
<td>7.897</td>
</tr>
<tr>
<td>Central NC</td>
<td>-0.497</td>
<td>5174</td>
<td>6.341</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>-1.166</td>
<td>11025</td>
<td>6.834</td>
</tr>
<tr>
<td>Western NC</td>
<td>-1.509</td>
<td>1861</td>
<td>6.492</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENDER + ETHNICITY</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female + Lumbee</td>
<td>1.339</td>
<td>305</td>
<td>9.407</td>
</tr>
<tr>
<td>Female + Black</td>
<td>0.723</td>
<td>12791</td>
<td>7.035</td>
</tr>
<tr>
<td>Male + White</td>
<td>0.457</td>
<td>2633</td>
<td>7.788</td>
</tr>
<tr>
<td>Female + White</td>
<td>0.043</td>
<td>985</td>
<td>7.708</td>
</tr>
<tr>
<td>Male + Lumbee</td>
<td>0.044</td>
<td>439</td>
<td>4.924</td>
</tr>
<tr>
<td>Male + Black</td>
<td>-0.252</td>
<td>2131</td>
<td>6.538</td>
</tr>
<tr>
<td>Male + Latino</td>
<td>-1.059</td>
<td>1540</td>
<td>6.573</td>
</tr>
<tr>
<td>Female + Latina</td>
<td>-1.188</td>
<td>3049</td>
<td>5.785</td>
</tr>
</tbody>
</table>

SPEAKER (random effect)

| Std. Dev. | 1.056 |

Model

<table>
<thead>
<tr>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>142,998.7</td>
<td>16</td>
<td>7.227</td>
<td>6.959</td>
<td>23,873</td>
</tr>
</tbody>
</table>

Not selected as significant: AGE, YEAR OF BIRTH.

In sum, it appears that region really is quite important for speech rate, but primarily in the indirect sense of effecting utterance length, which in turn has a hugely significant ($p = 0$) linguistic effect on speech rate. On the other hand, the gender + ethnicity category appears to effect speech rate differently than it does utterance length. I leave further consideration of this to future work.
impacted by utterance length. This process – commonly referred to as “anticipatory shortening” – is likely a phonetic universal, whereby longer utterances are spoken at a higher rate due to physiological (namely, respiration) limitations. All in all, these findings support Quené’s conclusions that “differences among regions … may be significant, but these differences are not robust” (2008: 1111) and, more broadly, that social variation in speech rate is mitigated by the strong effect of utterance length.

Nonetheless, we close this chapter noting that the findings here, with respect to social factors, confirm that there is social variation in speech rate, even if it is a mitigated confirmation. We now turn to the ways that speech rate and pause may be related.
7. The relationship between speech rate and pause

7.1. Are speech rate and pause related?

It is more than intuitive to think that speech rate and pause duration may be related. As mentioned above, Goldman-Eisler (1961) – and others (e.g., Kowal and O’Connell 1980) – found that perception of speech rate was strongly influenced by the amount and duration of pauses in the speech stream. Meanwhile, many measures of speech rate (e.g. the *speaking rate* measure, see §6.2) even include short pauses, explicitly and inextricably linking the rate of speech to pause.

While I did not draw attention to this fact in Chapters 5 and 6, the mixed-effect models presented for pause (Table 5.4.4) and speech rate (Table 6.5.4 & Table 6.5.5) had similar orderings for the region factors, though inverse. Namely, the pause model found Washington, DC and Western NC speakers to most favor long pauses, while the speech rate models found these two region-factors to favor slower speech rates (when utterance length was not included as a factor). In terms of ethnicity, in the pause model Lumbees and Latinos were found to favor longer pauses, while in the speech rate models they favored slower speech rates. These empirical findings on their own point to a relationship between pause and speech rate in need of further examination.

In this chapter, I assess the influence of pause duration on speech rate and vice-versa. As in Chapter 6, speech rate here refers to a measure of *articulation rate*, the rate
in terms of syllables per second of actual phonation produced by a speaker, not including silent pauses.\textsuperscript{1} Since this speech rate measure is pause-exclusive, it is methodologically independent from the data obtained on pause. Any correlations between pause duration and speech rate should be due to actual speech production effects.

\textbf{7.2. Modeling speech rate and pause duration}

In Chapters 5 and 6, I took advantage of mixed-effect modeling techniques to examine individual measurements of pause and speech rate. That approach had the advantage of allowing us to look for patterns across the entire large datasets, and, for example, to look at the effect of utterance length on speech rate on a per measurement basis. Here, I compress the data to examine summary data on the speakers as the individual tokens, focusing on their median values for speech rate and pause. That is, instead of multi-level data with pause and speech rate Ns of 22,734 and 23,871 respectively, we here examine a dataset of 104 tokens, one for each speaker. The data used here are combined from Tables 5.4.1 and 6.5.1. As mentioned before, the complete data for the 104 speakers are presented in the Appendix.

\textsuperscript{1} While I continue to focus on silent pauses – and not filled pauses (and, as described in §6.4, filled pauses are not included in the speech rate measurements here) – it is also worth noting that there is surely a relationship between filled pauses and speech rate. Filled pauses, such as \textit{uh} and \textit{um}, are often produced as lengthened syllables and therefore we might assume slow the rate of speech in comparison to a speaker’s “normative” rate.
Figure 7.2.1. Speech rate and pause by ethnicity and gender, with best-fit lines.
Figure 7.2.1 displays plots grouped by ethnicity and gender showing the relationship between speakers’ speech rates and pause durations. For all groups other than the Lumbee speakers, there is practically no direct correlation between speech rate and pause duration ($r^2 < 0.25$). For the Lumbees however, there is a quite strong correlation for both genders – with pause duration decreasing as speech rate increases ($r^2 = 0.95$ for both genders). An inverse linear relationship makes sense for speech rate and pause – and seems congruent with my earlier observations based on the comparison of the independent mixed models – but the pattern begs the question: Why only among the Lumbees’ data? Since all the Lumbees come from the same region, we have to ask whether this result is an outcome of their shared region, a possibility that is somewhat masked by the plots in Figure 7.2.1 (despite the shape of each datum which corresponds to region).

Submitting these data for the response variable of median speech rate to a fixed-effect model analysis, again using Rbrul, a step-down analysis yields the model presented in Table 7.2.1. This model indicates that region is, in fact, the strongest predictor ($p < 0.01$). We also find that the second most significant factor is a speaker’s

---

2 We will see this inverse relationship pattern again in Chapter 9, when we examine interactional effects on speech rate and pause and intra-speaker variation.
3 Since the data are collapsed to one token per speaker, the model no longer needs to account for the random effect of the individual speakers. It is thus a fixed-effect model, and not a mixed-effect model.
4 A step-up fixed-model analysis finds only region ($p < 0.001$) and ethnicity ($p < 0.05$) to influence speech rate. I believe that this mismatch is due to the fact that the step-up model has a slightly higher deviance, of 38.016, compared to the deviance of 35.234 for the step-down model (indicating a worse model) but a lower number of degrees of freedom, df = 10, compared to df = 12 of the step-down model (indicating a simpler model). In this analysis I opt to follow the slightly more complex, but better fitting, step-down model.
median pause duration\(^5\) \((p < 0.05)\). Although the strength of the effect of pause duration appears small with a log-odds value of -0.002, note that that is in terms of a +1 ms change in pause duration, so that a +100 ms change in pause would have a log-odds value of -0.200 – a fairly considerable amount. Gender and ethnicity are also found to influence speech rate \((p < 0.05\) for each).

**Table 7.2.1. Fixed-effect model for influences on speech rate**

<table>
<thead>
<tr>
<th>REGION</th>
<th>(p = 0.00333)</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texas</td>
<td>0.394</td>
<td>24</td>
<td>5.126</td>
<td></td>
</tr>
<tr>
<td>Southern NC</td>
<td>0.343</td>
<td>19</td>
<td>5.041</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>0.094</td>
<td>7</td>
<td>5.229</td>
<td></td>
</tr>
<tr>
<td>Eastern NC</td>
<td>0.067</td>
<td>5</td>
<td>5.018</td>
<td></td>
</tr>
<tr>
<td>Washington, DC</td>
<td>-0.131</td>
<td>12</td>
<td>4.431</td>
<td></td>
</tr>
<tr>
<td>Central NC</td>
<td>-0.271</td>
<td>29</td>
<td>4.518</td>
<td></td>
</tr>
<tr>
<td>Western NC</td>
<td>-0.496</td>
<td>8</td>
<td>4.220</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MEDIAN PAUSE DUR</th>
<th>(p = 0.0297)</th>
<th>Log-odds</th>
<th>(continuous predictor)</th>
<th>+1</th>
<th>-0.002</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>GENDER</th>
<th>(p = 0.0397)</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.141</td>
<td>47</td>
<td>4.994</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.141</td>
<td>57</td>
<td>4.627</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ETHNICITY</th>
<th>(p = 0.0419)</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0.315</td>
<td>31</td>
<td>5.151</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.047</td>
<td>34</td>
<td>4.594</td>
<td></td>
</tr>
<tr>
<td>Latino</td>
<td>-0.102</td>
<td>29</td>
<td>4.629</td>
<td></td>
</tr>
<tr>
<td>Lumbee</td>
<td>-0.260</td>
<td>10</td>
<td>4.831</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35.234</td>
<td>12</td>
<td>5.324</td>
<td>4.793</td>
<td>104</td>
</tr>
</tbody>
</table>

Not selected as significant: AGE, YEAR OF BIRTH, SPEECH RATE \(N\), PAUSE \(N\).

\(^5\) As in Chapter 5, pause duration is included in the model in raw, non-log units. This does not impact the model in any significant way and makes interpreting the results more intuitive.
Age and year of birth are found to be non-significant factors, as are a speaker’s number of pause measurements and speech rate measurements.\footnote{The speech rate and pause \textit{Ns} were included in order to test whether there was a correlation between the number of measurements and the outcome speech rate median or pause median. A correlation along these lines would not be good. It would indicate that there was a relationship between speakers’ central tendencies and the number of tokens available of them. Happily, there is no such relationship in these data.}

This model does not resolve whether the inverse relationship observable for the Lumbee speakers (in Figure 7.2.1) is related to region. In order to assess whether the lack of direct linear correlations for the other speakers may be a result of a potential region + ethnicity (+ gender) interaction, I plot in Figure 7.2.2 a few sub-samples of the other speakers, limiting ethnicities by region – African Americans from Washington, DC (both genders, $r^2 = 0.20$), European Americans from Texas (both genders, $r^2 = 0.01$), Latinos and Latinas from Texas ($r^2 = 0.02$), Lumbees (both genders, $r^2 = 0.90$), European Americans from Southern NC (the same region as the Lumbees; both genders, $r^2 = 0.09$), and African Americans from Southern NC (the same region as the Lumbees; both genders, $r^2 \approx 0$). In the figure, and through the poor correlation coefficients for all but the Lumbees, we see there is not a clear inverse relationship between pause and speech rate for any group other than the Lumbees, not even for the other ethnic groups from the same location.
Examining the same data from the opposite perspective – with speakers’ median pause durations as the response variable – obtains the fixed-effect model presented in Table 7.2.2. Importantly, this is the one case out of the many (many) models I ran where examining log-median pause durations (in log-ms) instead of raw median pause durations
(in ms) had an impact on the outcome of the model. So in this model the response variable is log-ms of pause duration.\(^7\)

**Table 7.2.2. Fixed-effect model for influences on pause log-duration**

<table>
<thead>
<tr>
<th>REGION</th>
<th>(p = 0.00121)</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washington, DC</td>
<td></td>
<td>0.221</td>
<td>12</td>
<td>6.147</td>
</tr>
<tr>
<td>Texas</td>
<td></td>
<td>0.102</td>
<td>24</td>
<td>6.028</td>
</tr>
<tr>
<td>Eastern NC</td>
<td>-0.002</td>
<td></td>
<td>5</td>
<td>5.977</td>
</tr>
<tr>
<td>Southern NC</td>
<td>-0.003</td>
<td></td>
<td>19</td>
<td>5.933</td>
</tr>
<tr>
<td>Western NC</td>
<td>-0.034</td>
<td></td>
<td>8</td>
<td>5.920</td>
</tr>
<tr>
<td>Central NC</td>
<td>-0.035</td>
<td></td>
<td>29</td>
<td>5.928</td>
</tr>
<tr>
<td>Ohio</td>
<td>-0.249</td>
<td></td>
<td>7</td>
<td>5.680</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENDER</th>
<th>(p = 0.00181)</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.076</td>
<td></td>
<td>47</td>
<td>6.014</td>
</tr>
<tr>
<td>Female</td>
<td>-0.076</td>
<td></td>
<td>57</td>
<td>5.919</td>
</tr>
</tbody>
</table>

**MEDIAN SPEECH RATE**

<table>
<thead>
<tr>
<th>(p = 0.0441)</th>
<th>Log-odds</th>
<th>(continuous predictor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>-0.073</td>
<td></td>
</tr>
</tbody>
</table>

**Model**

<table>
<thead>
<tr>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.915</td>
<td>9</td>
<td>6.301</td>
<td>5.962</td>
<td>104</td>
</tr>
</tbody>
</table>

Not selected as significant: ETHNICITY, AGE, YEAR OF BIRTH, SPEECH RATE \(N\), PAUSE \(N\).

This model indicates that region and gender are both almost equivalent predictors of pause duration (at the \(p < 0.01\) level). Speech rate is also found to be a predictor (\(p <

---

\(^7\) Using the raw pause duration medians as the response variable obtains only region and gender as significant. In that model, the \(p\) values for region and gender are slightly higher though (\(p = 0.002\) and \(p = 0.004\), respectively) than they are in the log-pause model. I suspect that the \(p\) value for speech rate in the non-log model is just over the \(p < 0.05\) threshold, since it is so close to that threshold in the log-pause model.

162
0.05), with increases in speech rate leading to small decreases (log-odds of -0.073) on log-pause duration.\(^8\) We note that, in this model, ethnicity is not found to be significant.

In conclusion, this analysis appears to support the general observation that there is an inverse relationship between pause duration and speech rate, despite the fact that this relation does not appear as a simple linear correlation \((r^2 = 0.02\) for all speakers\) for any group other than the Lumbees. This exception of the Lumbees, however, raises an interesting question, which we turn to here in further depth.

7.2.1. Why are the Lumbees different?

The above has not satisfactorily addressed why the Lumbees in particular show such a strong correlation between speech rate and pause, with no gender effect, while for the other groups this pattern is less clear (though obtains through the regression analysis). At this point, I can only speculate, but I do so here because I think the possible explanations are provocative.

Of all the “groups” I examine through the 104 speakers analyzed here, the Lumbees are most “group-like” of the speakers. That is, they are the most endocentric,

---

\(^8\) Again, it is important to keep the difference between units of speech rate and pause duration in mind. Pause, as in the model presented in Table 7.2.1, was measured in milliseconds, so a +1 increase, or -1 decrease, was a very minor change – differences on order of +100 ms and -100 ms are likely between two pause measurements (the standard deviation for pause duration for all the speakers is 101 ms). Speech rate, on the other hand, is measured in \(\sigma/\text{sec}\) with a +1 or -1 difference between two measurements being a fairly large difference (the standard deviation for speech rate for all speakers is 0.72 \(\sigma/\text{sec}\) ). This is to say that the larger log-odds value for speech rate in the pause model of Table 7.2.2 actually more likely represents a less strong effect than the very small log-odds value for the effect of pause on the speech rate model of Table 7.2.1.
sharing more than ethnic and regional affiliation (as the Washington, DC African Americans and the Texas Latinos do). The Lumbees are culturally unified (cf. Wolfram, Dannenberg, Knick, and Oxendine 2002) and, thus, we might expect them to share norms more so than other divisions of the speakers based on simplistic ethnicity, region, and gender grounds. It seems plausible that the sort of direct relationship between pause and speech rate could be found only in the tightest knit speech communities.

To humor this digression for one more moment, this speculation may have interesting bearings on broader sociolinguistic theory. We could consider, for example, all of the Washington, DC speakers to comprise a community of practice (Holmes and Meyerhoff 1999; Eckert 2000) through their shared participation in a summer program (not to mention their similar ages, ethnicities, and socioeconomic and educational backgrounds; cf. §9.5; Mallinson and Kendall 2009). With the exception of the Lumbees, these Washington, DC speakers do have a higher degree of correlation ($r^2 = 0.20$) than any other group in Figure 7.2.2. Yet, it is still a very low correlation.

As I mentioned above, the Lumbees can be considered to comprise a “group” more so than the other speakers examined and the notion of speech community seems more simplistically (cf. Patrick 2002) applied to them than other subsets of the speakers used in this study. Much recent sociolinguistic work has explored the ways in which more nuanced understandings of social practices – such as the community of practice concept – explain language variation (e.g., Holmes and Meyerhoff 1999; Eckert 2000; Mallinson 2006; Mallinson and Childs 2007), but the evidence here points more to “old-fashioned” notions of speech community as perhaps coming into play.
Less speculatively, and perhaps more importantly, a number of studies have been undertaken which have shown the degrees to which Lumbee English displays characteristics of other Native American English varieties or otherwise differentiates itself from other regional and standard varieties of American English (Wolfram and Dannenberg 1999; Torbert 2001; Coggshall 2008). Accordingly, the finding that Lumbees favor the longest pauses and favor slower speech rate seems in line with these previous descriptions of Lumbee English and research on other Native American language varieties (e.g., Philips 1976). While we need to examine a larger number of Lumbee speakers to fully understand this, we can putatively interpret this pattern in Lumbee speech as yet another, albeit subtle, way in which the Lumbees reflect their Native American heritage through their English dialect (cf. Wolfram et al. 2002).

7.3. In closing: Variation in speech rate and pause duration

As this chapter has shown, there is an, albeit small, inverse correlation between speech rate and pause. This correlation is quite striking for the Lumbee speakers, but is more subtle for the majority of the data, emerging through the regression analyses of Tables 7.2.1 and 7.2.2. Importantly, as I pointed out earlier, since speech rate, as analyzed here, is pause-exclusive, there is nothing mechanistic that connects these two measures. Instead, what we see here is the complex coordination of cognitive, physiological (i.e. articulatory and respiratory), and discourse processes. Despite this correlation, which we might suppose is somewhat of a universal along all three
dimensions (cognitive, physiological, and discourse-related), we also persistently find social variation – pause and speech rate realizations that are at least somewhat determined by a speakers’ region, gender, and/or ethnicity.

Finally, it is also worth noting that for all of the models that found region to be significant in this and the previous two chapters (that is, all of the models other than the mixed-effect model for speech rate, Table 6.5.6, which included utterance length as a factor), region was found to be the most significant factor. It consistently trumped gender and ethnicity and whatever other non-linguistic (that is, non-utterance length) factors that were found to be significant.

The results here, I believe, do not indicate that there is something essential about the patterning of speech rate and pause in terms of region – that is, I do not think that there is something necessarily special about Ohio or Ohioan English that causes those speakers to have shorter pauses and higher speech rates than most of the speakers of the other regions. Instead, I think that the results, as far as the category of “region” is concerned, indicate that pause duration and speech rate, are – in addition to the outcomes and indicators of cognitive processes – the results of both sociohistorical processes and localized practices, as are so many aspects of dialect.

In other words, for the data examined here, with the exception of the Lumbees – who do pattern together primarily on ethnic (+ regional) grounds – it appears that the variation is primarily differentiated by region, more so than gender or ethnicity. This, I believe, has important consequences for an understanding, however putative, of the influence of social factors, and socialization, on features such as pause and speech rate. It
indicates that speakers’ timing features – such as pause and speech rate – are likely influenced by the matrix talk of speakers’ linguistic development; that is, that pausing and speech rate are reflective of subtle (or covert in the terms of language change research; cf. Labov 2001) influences on speech patterns more than they are available (or overt) as indices of ethnic affiliation or as features that can be made use of for identity work.

Whether these regional differences are in fact the outcomes of sociohistorical processes or whether they are even more localized than that – a sort of micro-regional differentiation – must remain a question for the future.
8. Style and register variation in speech rate and pause duration

8.1. Introduction

The previous chapters have focused on large-scale analyses of the speech rate and pause data derived from the SLAAP archive. In order to assess overarching patterns in the data, I have downplayed potential qualitative differences between the interviews that I examined. For example, the Washington, DC data come from sociological interviews as opposed to sociolinguistic interviews (Mallinson 2007; Mallinson and Kendall forthcoming; cf. §9.5). And it is plausible that these sorts of differences impact the speakers’ pause and speech rates. In this chapter, I consider register and speech style as potential correlates of pause duration and speech rate.

It seems intuitive that pause and speech rate are stylistically impacted in the Labovian sense of attention paid to speech (Labov 1966, 1972). In fact, pause and speech rate are used by Labov (more in theory however than in practice) as paralinguistic cues for determining the categorization of style – e.g., as careful or casual – for a particular stretch of speech. In Chapter 10, I return to the notion of pause and speech rate as paralinguistic cues and take up the question of how they relate to features such as phonological and morphosyntactic variables. In this chapter, I examine principally the role of style – mostly operationalized here in terms of register – as an independent variable on speakers’ speech rates and pause durations. I also, briefly, assess the degrees
to which changes in speech rate and pause correlate with changes in the realization of “normal” sociolinguistic variables.

8.2. Pause and speech rate and stylistic variation

Kendall and Wolfram (forthcoming) analyzed the speech of three local leaders in small African American communities in North Carolina in different speech situations in order to examine style-shifting and its relationship to local vernacular features and norms. In terms of morphosyntactic and phonological factors, they found very little evidence that the speakers they examined shifted away from their vernacular in the more formal, or public, contexts. I here revisit Kendall and Wolfram’s (forthcoming) work focusing on their pause and speech rate data and considering those data in relation to their morphosyntactic and phonological data.¹

8.2.1. Three case studies

In Table 8.2.1 and Figure 8.2.1,² I present data from two speech situations on seven morphosyntactic and phonological variables, along with median pause duration and

---

¹ In addition to Kendall and Wolfram (forthcoming), some of the phonological and morphosyntactic data used in this chapter come from Rowe and Kendall (2004), Carpenter (2004, 2005), D’Andrea (2005), Rowe (2005), Carpenter and Hilliard (2005), and Vadnais (2006).

² To enhance readability, in Figures 8.2.1, 8.2.2, and 8.2.3, median pause duration is presented in centiseconds (hundredths of a second, not milliseconds) and median speech rate has been multiplied by 10.
speech rate measures, for a county commissioner from Roanoke Island in Dare County, North Carolina, the “Commissioner” of the Appendix (one of the Eastern NC speakers). Kendall and Wolfram (forthcoming) examined her speech in a public address in relation to her speech in a sociolinguistic interview.

Table 8.2.1. Comparison of pause and speech rate with diagnostic variables in two speech settings for the County Commissioner from Roanoke Island, NC

<table>
<thead>
<tr>
<th></th>
<th>Static loc. to</th>
<th>Copula Abs</th>
<th>3rd sg. -s Abs</th>
<th>Past tense be Reg.</th>
<th>Pre-V CCR</th>
<th>Post-V rlessness</th>
<th>Neg. Concord</th>
<th>Median Speech Rate</th>
<th>Median Pause Dur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Address</td>
<td>1/8</td>
<td>1/11</td>
<td>0/7</td>
<td>3/41</td>
<td>15/46</td>
<td>4/141</td>
<td>0/1</td>
<td>3.35 s/s</td>
<td>454 ms</td>
</tr>
<tr>
<td>12.5%</td>
<td>9.1%</td>
<td>0.0%</td>
<td>7.3%</td>
<td>32.6%</td>
<td>2.8%</td>
<td>0.0%</td>
<td></td>
<td>3.92 s/s</td>
<td>371 ms</td>
</tr>
<tr>
<td>Socioling. Interview</td>
<td>2/11</td>
<td>0/27</td>
<td>1/18</td>
<td>4/35</td>
<td>10/47</td>
<td>6/171</td>
<td>7/12</td>
<td>3.35 s/s</td>
<td>454 ms</td>
</tr>
<tr>
<td>18.2%</td>
<td>0.0%</td>
<td>5.6%</td>
<td>11.4%</td>
<td>21.3%</td>
<td>3.5%</td>
<td>58.3%</td>
<td></td>
<td>3.92 s/s</td>
<td>371 ms</td>
</tr>
</tbody>
</table>

Figure 8.2.1. Graphic comparison of selected variables from Table 8.2.1

In later chapters I will also use this transformation of the speech rate and pause data whenever I need to portray both features on the same plot or numerical scale.
We did not find significant variation among the morphosyntactic and phonological variables depending on the speech event ($\chi^2 = 0.11$, $p > 0.05$). T-tests for pause and speech rate indicate that the Commissioner’s pause realizations, despite having markedly different medians, are not significantly different ($p > 0.05$), but that her speech rates are significantly different across speech situations ($p < 0.001$).

In Table 8.2.2 and Figure 8.2.2, I present six morphosyntactic and phonological variables for the Town Manager of Princeville, NC, the “Town Manager” of the Appendix (a Central NC speaker). In addition to his sociolinguistic interview, we have an interview that was recorded with him and broadcast on the statewide NPR program “The State of Things”. As with the Commissioner, Kendall and Wolfram (forthcoming) do not find significant differences between the two contexts for his radio interview and sociolinguistic interview in terms of his morphosyntax and phonology ($\chi^2 = 0.40$, $p > 0.05$).

<table>
<thead>
<tr>
<th></th>
<th>Median Speech Rate</th>
<th>Median Pause Dur.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Radio Interview</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plural –s Abs</td>
<td>5/24</td>
<td>5.34 s/s</td>
</tr>
<tr>
<td>Copula Abs</td>
<td>5/9 (4/7 are)</td>
<td>300 ms</td>
</tr>
<tr>
<td>3rd sg. –s Abs</td>
<td>10/16</td>
<td></td>
</tr>
<tr>
<td>Past tense be Reg.</td>
<td>5/7</td>
<td></td>
</tr>
<tr>
<td>Pre-V CCR</td>
<td>5/11</td>
<td></td>
</tr>
<tr>
<td>Post-Vocalic r-lessness</td>
<td>42/68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16.7%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55.6%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>71.4%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>61.7%</td>
<td></td>
</tr>
<tr>
<td><strong>Socioling. Interview</strong></td>
<td>19/96</td>
<td>5.27 s/s</td>
</tr>
<tr>
<td>Plural –s Abs</td>
<td>7/31 (3/7 are)</td>
<td>322 ms</td>
</tr>
<tr>
<td>Copula Abs</td>
<td>6/10</td>
<td></td>
</tr>
<tr>
<td>3rd sg. –s Abs</td>
<td>26/36</td>
<td></td>
</tr>
<tr>
<td>Past tense be Reg.</td>
<td>30/57</td>
<td></td>
</tr>
<tr>
<td>Pre-V CCR</td>
<td>46/100</td>
<td></td>
</tr>
<tr>
<td>Post-Vocalic r-lessness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.8%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22.6%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72.2%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52.6%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>46%</td>
<td></td>
</tr>
</tbody>
</table>

3 Following Kendall and Wolfram (forthcoming), the chi-squared calculations in this chapter exclude r-lessness in all cases.
We do notice that he has a slightly shorter median pause in his radio interview, but that for all intents and purposes his pause realization and speech rate are in line with findings for his morphosyntactic and phonological variables. He doesn’t style-shift to any significant extent despite the different situational contexts and our presumption that these two settings might yield different registers (t-tests yield $p > 0.05$ for both speech rate and pause).

Finally, in Table 8.2.3 and Figure 8.2.3, I present a similar table and graphic for the (at the time of the recordings) Mayor of Princeville, the “Mayor” of the Appendix (a Central NC speaker).

Table 8.2.3. Comparison of pause and speech rate with diagnostic variables in two speech settings for the Mayor of Princeville, NC

<table>
<thead>
<tr>
<th></th>
<th>Plural −s Abs</th>
<th>Copula Abs</th>
<th>3rd sg. −s Abs</th>
<th>Past tense <em>be</em> Reg.</th>
<th>Pre-V CCR</th>
<th>Post-Vocalic <em>r</em>-lessness</th>
<th>Median Speech Rate</th>
<th>Median Pause Dur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech at Town Event</td>
<td>14/28</td>
<td>6/13 (4/6 are)</td>
<td>10/18</td>
<td>0/0</td>
<td>79</td>
<td>39/93</td>
<td>4.65 σ/s</td>
<td>500 ms</td>
</tr>
<tr>
<td>Socioling. Interview</td>
<td>30/94</td>
<td>7/12 (4/4 are)</td>
<td>17/18</td>
<td>15/24</td>
<td>45/66</td>
<td>51/100</td>
<td>4.74 σ/s</td>
<td>334 ms</td>
</tr>
</tbody>
</table>
In addition to the sociolinguistic interview, the NCLLP field workers were able to record the Mayor giving a public speech in honor of the town’s “birthday”. Despite the fact that Kendall and Wolfram (forthcoming) find no statistically significant difference between the contexts for her morphosyntactic and phonological variables ($\chi^2 = 0.01$, $p > 0.05$) and her speech rates are noticeably similar (a t-test finds $p > 0.05$), we do notice here a significant difference between her pause duration in the sociolinguistic interview and in her public speech, which is confirmed by a t-test ($p < 0.001$).

![Figure 8.2.3. Graphic comparison of variables from Table 8.2.3](image)

8.2.2. *Speech rate and pause across two settings for the three speakers*

How do we account for the differences among these three speakers? For each speaker, we might expect that the two settings would predict (or require) different registers and/or different stylistic presentations. None of the speakers show a significant
mono-directional shift in their morphosyntactic or phonological variables. Yet each of the three speakers realizes a different possible option across their two settings in terms of pause and speech rate: The Town Manager exhibited no significant variation in pause and speech rate between a sociolinguistic interview and a public radio interview; the Commissioner showed significantly slower speech rate (of about a half a syllable per second) in a public address than in her sociolinguistic interview; and, finally, the Mayor showed significantly longer pauses during her public speech than in her sociolinguistic interview.

![Figure 8.2.4. Changes in speech rate and pause duration for three speakers](image)

Figure 8.2.4. Changes in speech rate and pause duration for three speakers

---

4 The difference between the Commissioner’s speech rates is only about a half a syllable per second (~15%), but note that this is much larger than the JND of 5% (Quené 2007). See §5.6 and §6.3.
As illustrated in Figure 8.2.4, even though many of the differences in these speakers’ pause durations and speech rates are not statistically significant, it is worth noting that for all three speakers, including the Town Manager, we see an inverse relationship between changes in speech rate and pause. At the same time, the Manager shows the opposite pattern from the other two speakers: His sociolinguistic interview setting obtains longer pauses and slower speech rates than his radio interview. For the other two speakers, the sociolinguistic interview has the shorter median pause duration and faster speech rate.\(^5\)

For the Mayor, it appears possible that she read her speech, although even if she did not, the situation called for a “read-like” register and the mayor responded that way, with significantly longer pauses than she used in her sociolinguistic interview. This may be true for the Commissioner as well. The Commissioner gave her public speech over the backdrop of a guitar soloist. Whether or not her speech was read, she perhaps adopted the cadence of the guitar playing and this influenced her speech rate. The Town Manager, on the other hand, although surely aware of the state-wide audience of his public radio interview, clearly was not reading anything. It is further unlikely that his talk was planned whatsoever, whereas, even if the Mayor and Commissioner did not “read” their speeches, they clearly had pre-planned them, and this, perhaps, accounts for

\(^5\) If, following Labov (1966, 1972) and Chapter 10, we consider pause and speech rate as paralinguistic cues to attention to speech and we assume that longer pauses and slower speech rates are indicative of higher attentiveness or formality on the part of the speaker, then we can make the interesting observation from Figure 8.2.4 that the Manager appears to treat his sociolinguistic interview as a more formal or self-conscious speech situation than his radio interview.
the lack of significant differences for the Manager’s speech rate and pause in relation to the Commissioner and Mayor.

We see here that pause and speech rate can be – but are not categorically – impacted by style and register.6 This does not feel like a groundbreaking observation. In fact, it might be tempting to say that the finding that pause and speech rate realizations correspond with a reading-like style is so explainable, understandable, and in line with previous research (see §5.2 and §6.3) that it is not interesting, but I think it does warrant consideration. Reading passage and public speech are common register types,7 so the fact that pause realization and speech rate correspond with style/register can be seen as an important step towards understanding pause and speech rate as variable features of speech, which correlate with social and stylistic variables. As we can clearly see from the “traditional” or “canonical” variables presented in Tables 8.2.1 – 8.2.3, variable realizations can vary in ways that are not aligned or parallel with one another.

8.3. In closing: Register and style variation in pause and speech rate

This short chapter has supported the notion that speech rate and pause can vary according to speech style, especially at the level of register. Furthermore, the short case

6 Interestingly, some of the psycholinguistic literature reviewed in Chapters 5 and 6 indicates that speech rate should increase and pauses should shorten with planning (e.g., Goldman-Eisler 1961, 1968). However, we see in both cases where there is a significant change – the Mayor’s pause duration and the Commissioner’s speech rate – that the change is towards a lengthening of the pause time and a slowing of the speech rate. These speakers appear to (optionally, at least) show stylistic shift towards “read-like” style more than they appear to show evidence of pre-planning in their talk.

7 In Labovian terms, reading passage is a style in the attention to speech model and seems also appropriate to consider public speech as an attention to speech style.
studies examined in this chapter (and Kendall and Wolfram forthcoming) seem to indicate that pause and speech rate may perhaps best be viewed as impacted at the discourse or interactional level: Don’t speakers make use of pause and speech rate, consciously or subconsciously, for stylistic effects and for presenting themselves in certain ways in relation to their interlocutors? It is clear from the putative analysis here that straightforward quantitative analyses cannot alone provide the answers. In the next chapter, we’ll complicate this analysis of pause and speech rate by investigating the effects of interlocutors on pause and speech rate realization – we will follow the hypothesis that these features might be productively approached as interactional variables.
9. Pause and speech rate in interaction

9.1. Introduction

In the last chapter we confirmed that style plays a role – albeit not a mono-directional role – in the realization of pause and speech rate. In this chapter, we will take a similar, though deeper, look into pause and speech rate. Instead of considering these speech features as attributes of individuals or social groups, or components of individuals’ speech in relation to particular registers or levels of attention to speech, we will examine these features more holistically as arising in the speech of individuals within particular, multi-dimensional interactional events. In other words, we will take the next step from the discussion of style in the previous chapter, to considering something more nuanced. This view of style as “interaction” might be best considered within recent conceptions of style (cf. Eckert and Rickford 2001; Schilling-Estes 2002; Eckert 2005; Coupland 2007).¹

9.2. Some problems with Chapter 5, 6, and 7

¹ While my goals are broader, the approach to style operationalized here could be argued to be more appropriately considered as an implementation of audience design (Bell 1984, 2001).
The evidence presented in Chapters 5 through 7 indicated a number of patterns with respect to speech rate and pause for the regional, ethnic, and gender dialects of the speakers in the study. Readers may have noted, however, that my claims were narrow. I did not make assertions about the speech rates of the ethnic dialects of African American English or Latino English or European American English in monolithic terms. In fact, I was careful to keep clear that the findings were relevant for the samples examined, instead of arguing that they were indicative of general differences between the speech of African Americans, Latinos, and European Americans. Also, while, for example, the fixed-effect model for the role of speech rate on (log-)pause duration (Table 7.2.2) indicated that Ohioans have the shortest pauses, I did not use this finding to proclaim that – indeed – Northerners talk faster than Southerners (since many findings have shown pause duration to influence perception of rate of speaking more than actual speech rate; see §5.2).

My hesitancy to emphasize the substantive findings of the earlier analyses can be read as a necessary hedge, partly due to the “convenience” sampling method use here. That is, the sample used here is based on what is currently available in the SLAAP archive rather than on a more fully representative and balanced corpus. At the same time, however, my restraint in claiming global significance for the findings so far stems primary from my belief that the full range of variation within the pause and speech rate data is not fully captured by the analyses of the previous chapters. In fact, it seems clear – from the literature review presented in Chapters 5 and 6, from the analysis of stylistic variation for the Mayor, the Town Manager, and the Commissioner in Chapter 8, and
even from our commonsense intuitions and observations about speaking patterns – that speech features like pause and speech rate are impacted by a larger range of *interactional* features than have so far been considered. To consider this more deeply, in this chapter we approach pause and speech rate as, what we might term, *interactional variables*.

Returning to the notion of style, the fairly straightforward *attention to speech* model – as simplistically employed, for example, in the previous chapter – does not seem sufficient for fully understanding the data. Other approaches to style, such as *audience design* (Bell 1984, 2001) or *speaker design* (cf. Schilling-Estes 2002), seem more compatible with the notion of interactional variation. However, it is not clear how one can best operationalize approaches to style like audience design and speaker design quantitatively (cf. Renn and Terry 2007). To examine interactional variation, as I’m terming it, we will follow the strengths of the attention to speech model by maintaining a quantitative focus, but we will also incorporate advances made through the audience design and speaker design models by considering metrics that examine aspects of the speech events external to the speaker.

In an attempt to more holistically examine pause and speech rate, we now turn to a more complete quantitative analysis of speech rate and pause. First, we re-examine the dataset from Chapter 7, but we here account for possible interlocutor effects by considering the gender and ethnicity of the interviewers and the number of total interlocutors in each interaction. We then conduct two “case studies”, which take advantage of two sets of interviews in SLAAP where we have enough data to look at the differences that arise in individuals’ speech with different interlocutors.
9.3. Interlocutor effects on pause and speech rate

As explained in §4.2, most of the data examined thus far have been the speech of the interviewees in sociolinguistic interviews (and for Washington, DC sociological interviews), though I have also examined to a lesser degree the speech of the interviewers. We turn now to examine whether there are statistically significant effects on these interviewees’ pause and speech rates from the ethnicity and gender of their interviewer(s). There are two ways that we can frame this question, and we follow both framings. First, does it matter – in terms of median pause duration and median speech rate – whether the interviewers are of the same ethnicity and/or gender as the interviewee? Second, could the actual ethnicity and/or gender of the interviewer influence the median speech rate and pause duration of the interviewee, regardless of that interviewee’s ethnicity and/or gender? Further, we also ask here whether the number of participants in an interview quantitatively influences the speech rate and pause duration of the interviewees.

The examinations in this section deal with a subset of the data examined in Chapter 7. These are the overall median values for 96 of the speakers – all of the interviewees examined above with the exception of three, for whom the data came from changing sets of interviewers. In other words, I now examine 96 subjects who were interviewees with respect to their interviewer(s) and co-participants. To confirm that this subset of the data does not impact the comparability of the data here with the analyses
presented earlier, Table 9.3.1 compares the means of the 96 subjects to the means of the original 104 subjects by ethnicity (ordered alphabetically). Note that none of the differences are significant.

<p>| Table 9.3.1. Minor and non-significant differences between subset and main data |
| --- | --- | --- | --- | --- | --- | --- |</p>
<table>
<thead>
<tr>
<th>Subset N</th>
<th>Subset Mean Speech Rate</th>
<th>Overall Mean Speech Rate</th>
<th>Speech Rate T-test Results</th>
<th>Subset Mean Pause Duration</th>
<th>Overall Mean Pause Duration</th>
<th>Pause Duration T-test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>African Americans</td>
<td>31 (-3)</td>
<td>4.58 σ/sec</td>
<td>4.59 σ/sec</td>
<td>$p &gt; 0.05$</td>
<td>399 ms</td>
<td>399 ms</td>
</tr>
<tr>
<td>European Americans</td>
<td>27 (-4)</td>
<td>5.13 σ/sec</td>
<td>5.15 σ/sec</td>
<td>$p &gt; 0.05$</td>
<td>412 ms</td>
<td>405 ms</td>
</tr>
<tr>
<td>Latinos</td>
<td>29 (0)</td>
<td>4.63 σ/sec</td>
<td>4.63 σ/sec</td>
<td>$p = 1.0$</td>
<td>400 ms</td>
<td>400 ms</td>
</tr>
<tr>
<td>Lumbees</td>
<td>9 (-1)</td>
<td>4.89 σ/sec</td>
<td>4.83 σ/sec</td>
<td>$p &gt; 0.05$</td>
<td>389 ms</td>
<td>398 ms</td>
</tr>
</tbody>
</table>

9.3.1. The influence of gender and gender differences between speakers

In order to investigate whether the gender of a speaker’s interlocutors might influence their speech rate and pause, I coded each interviewee for the gender of their interviewer(s) and/or co-participant(s). These were marked as “female” if all the interviewers were female, “male” if all the interviewers were male, or “mixed” if the interviewers and co-participants were comprised of males and females – that is, if (other than the participant) there were both males and females taking part in the interaction. Of course, “mixed” interactions necessarily had at least two interviewers or co-participants in addition to the participant. Interactions marked as “male” or “female” had one, two, or
three participants in addition to the interviewee. I also coded whether the gender of the interviewee was the “same” or “different” than the gender(s) of their co-participant(s). Note that for “mixed” interviewers, this value was always coded as “different”.

For these gender data, the interesting patterns arise from a full consideration of the actual genders of the participants (more than the simpler “same” vs. “different” categorization). Figure 9.3.1 provides boxplots for the speech rate data, with each boxplot providing data for a pairing of the interviewers’ gender and the interviewee’s gender. It is immediately clear from this figure that speech rates are substantially lower for interviewees who were interviewed by only females.

![Boxplot of Interviewer Gender(s) and Participant Gender on Speech Rate](image)

**Figure 9.3.1. Boxplots of effect of interviewer and participant gender on speech rate**

A t-test finds the difference between the subjects interviewed by only females (mean = 4.21 σ/sec) and the remaining subjects (mean = 5.13 σ/sec) to be highly significant ($p < 0.0001$). The difference here, of 18%, is well above the 5% JND threshold indicating that these differences should, in fact, be quite perceivable to hearers (see §5.6 and §6.3;
Quené 2007). At the same time, the difference between the speech rates of the female subjects and the male subjects who were interviewed by females is not significant \( (p = 0.67) \), indicating that the important factor here is solely the gender of the interviewer, not the relationship between the interviewer’s and interviewee’s gender.

To better understand the effect of interviewer and co-participant gender on the subjects’ speech rates, we revisit the fixed-effect regression analysis from Table 7.2.1 of Chapter 7, but this time we add the new predictor of interviewer gender. Recall that that model found region, median pause duration, gender, and ethnicity to be significant predictors of a speaker’s (median) speech rate, in that order. This model, found from a step-up/step-down analysis using Rbrul, is presented in Table 9.3.2.

**Table 9.3.2. Fixed-effect model for influences on speech rate, including interviewer gender**

<table>
<thead>
<tr>
<th>INTERVIEWER GENDER</th>
<th>Log-odds</th>
<th>Tokens</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Interviewer(s)</td>
<td>0.343</td>
<td>43</td>
<td>5.155</td>
</tr>
<tr>
<td>Mixed Interviewers</td>
<td>0.256</td>
<td>16</td>
<td>5.067</td>
</tr>
<tr>
<td>Female Interviewer(s)</td>
<td>-0.599</td>
<td>37</td>
<td>4.212</td>
</tr>
</tbody>
</table>

Model

Deviance: 26.624

Intercept: 4.812

Mean: 4.777

Total N: 96

Not selected as significant: REGION, ETHNICITY, GENDER, AGE, YEAR OF BIRTH, MEDIAN PAUSE DUR.

We find here a model strikingly different than those found in Chapter 7. The interviewer gender factor group is highly significant \( (p < 0.00001) \), with a log-odds range of 0.942. The ordering of the factors confirms the depiction of the data in the boxplot of Figure 9.3.1, female interviewers predict slower speech rates while mixed-gender interviewers
and male interviewers predict faster speech rates. Most importantly, with the gender of the interviewer(s) included as a factor, no other factor groups are now found to be significant.

Figure 9.3.2 provides similar boxplots for the pause duration data. The pattern found for speech rate does not appear evident here. The only visible difference among the data is that female subjects interviewed by males have shorter pauses (mean = 342 ms) than the other groups (mean = 417 ms). This difference is confirmed by a t-test ($p = 0.001$). It is hard to explain this pattern, however, and I will not attempt to do so here.

![Boxplot of Interviewer Gender(s) and Participant Gender on Pause Duration](image)

**Figure 9.3.2. Boxplots of effect of interviewer and participant gender on pause**

I do want to note, however, that the similarity between the remaining four groups – female interviewers with male participants, mixed gender interviewers, female interviewers with female participants, and male interviewers with male participants – seems both to confirm the general findings of the approach to pause used in Chapter 5 and to indicate that this exception to the conformity in the data (i.e. the female
participants interviewed by males stand out from the other pairings of interviewee-interviewer) is likely indicating something worthy of further investigation.

9.3.2. The influence of ethnicity and ethnicity differences between speakers

Having found that the interviewers’ gender may have striking effects on an interviewees’ speech rate, we now turn to the ethnicity of the interviewers and co-participants. As before with gender, I coded each participant in the sub-sample for the ethnicity/ies of their interviewer(s) and co-participants and for whether that ethnicity is different or the same as the interviewee’s.

Figure 9.3.3. Boxplots of the influence on speech rate of different/same ethnicity of interviewers and interviewees
Due to the larger spread of possible comparisons for ethnicity than was necessary for gender, I start by presenting in Figure 9.3.3 boxplots for speech rate and pause duration showing the influence on the aggregated speakers of whether they were interviewed by fieldworkers of their same ethnicity. For speech rate, we notice that speakers appear to speak more slowly when interviewed by someone of a different ethnicity (mean = 4.59 σ/sec vs. mean = 5.16 σ/sec), an observation confirmed by a t-test ($p < 0.0001$). Pause duration, on the other hand, is not observably or significantly different ($p > 0.05$).

Does this pattern for speech rate carry for all speakers, or only for speakers of a certain ethnicity? Figure 9.3.4 attempts to shed light on this question by separating the speech rate data by the ethnicity of the participant, and whether they are interviewed by someone of the same or different ethnicity.

![Boxplot of Participant Ethnicity v. Interviewer Ethnicity on Speech Rate](image.png)

**Figure 9.3.4.** Boxplots of speech rate based on participant ethnicity and the ethnicity of their interviewer(s)
The figure illustrates primarily that we do not have the data necessary to examine this phenomenon at length: There are no European Americans in the sample who were interviewed by non-European Americans; there is only one interview conducted by a Lumbee interviewer; and, there are no Latino/a interviewers whatsoever. However, Figure 9.3.4 does appear to show some evidence that African Americans have faster speech rates when interviewed by African Americans (mean = 4.44 σ/sec) than when interviewed by non-African Americans (mean = 5.51 σ/sec), but this difference is not found to be significant by way of a t-test ($p > 0.05$). It is possible that the large difference in Ns makes these two groups of speakers difficult to compare, since there are only four African American subjects interviews by an African American interviewer and 27 interviewed by non-African Americans. This pattern and putative finding is also mitigated by the fact that all four speakers interviewed by an African American fieldworker were from Southern NC and that the interviewer had previous relationships with his interviewees. In short, the familiarity of the interviewer and interviewees in these cases may have been more important than their ethnicities.

In closing this section, it seems clear that we need more diverse fieldworkers to better investigate this question, and other aspects of potential interviewer accommodation. If the pattern holds that the ethnicity of the interviewers impacts the speech of the interviewees, this finding – coupled with the finding of the influence of interviewer gender in the previous section – could have major repercussions on sociolinguistic field methods.
9.3.3. *The influence of the number of participants on speech rate and pause duration*

We now examine the effect of the number of overall participants in an interaction on a participant's (median) speech rate and pause duration. Figure 9.3.5 displays boxplots for speech rate and pause organized by the number of participants. Through these boxplots, we see a pattern whereby speech rate decreases as the number of participants increases from 2 to 4 and then sharply increases for interviewees in interactions with 5 participants (although we note that there are only four interviewees from interviews with 5 participants). This pattern is slightly evident for pause duration as well, but as in the previous two sections, we see less patterning of pause according to the interlocutors.

Figure 9.3.5. Boxplots of number of participants on speech rate and pause
The general pattern seen here, especially for speech rate, actually makes intuitive sense. While this pattern should be considered putative (since it is based on a low \( N \) for interactions with 5 participants), we can interpret this as showing that, for low numbers of speakers, speakers’ speech rates decrease as more interlocutors enter the conversation, as a sort of deferentiality or a strategy for ensuring that all hearers can follow one’s talk and all speakers have opportunities to share the floor. At a certain point, however, (as we see here, at 5 participants), the increased competition for talk-time created by the higher number of interlocutors may cause speakers to speak faster, to fit more contribution into their more limited talk-time. That is, with a large enough number of participants talk may become a competitive enterprise more than a collaborative enterprise. This is of course somewhat speculative, but I hope that it is at least provocative enough to inspire more work along these lines.

9.3.4. The effects of interviewers

All in all, the findings of this section raise some important questions about the effect of interviewers and co-participants on the speech obtained in sociolinguistic interviews (e.g., Hazen 2000). If, as this section finds with respect to speech rate, the realization of linguistic and sociolinguistic features may in some cases be primarily predicted by features of the interviewers, we may need to reconsider audience design (Bell 1984, 2001) as more central to the discipline, both theoretically and methodologically. The contribution of this section to our understanding of the impact of
the ethnicity of the interviewer(s) (and the ethnic differences between interviewers and interviewees) and the number of participants in an interaction are clearly limited due to the unbalanced nature of the data.

However, the distribution of interviewer gender is fairly evenly balanced and the data show a striking pattern when examined from the perspective of interview gender. I do not know how to interpret the fact that female interviewers elicit significantly slower speech from subjects or to account for the effect on the overall regression model of adding interviewer gender – of removing the significance of all other factor groups – other than to say this outcome may be one of the most consequential findings of this dissertation. This clearly needs further consideration.

We now move away from general, quantitative findings based to look more closely at two case studies, which we can approach with slightly deeper ethnographic insight.

9.4. A case study: Who is interviewing EH?

In 2007, Danica Cullinan, a member of the NCLLP and NCSU Masters student, organized five interviews with “EH,” an 82 year-old African American woman living in Raleigh, North Carolina, originally from Wilson County, North Carolina.² In these interviews, Cullinan set out to examine intra-speaker variation by changing the

² This section is based on work Danica Cullinan conducted for a project for English 584 at NCSU in the Fall semester of 2007. I am extremely grateful for her letting me use her data and draw on some of her analysis.
interviewer for each interview, while controlling for as many other factors as possible, including location, time of day, audio equipment, and “energy level of the interviewee” (Cullinan 2007: 6).

Table 9.4.1. Interviewers and data for EH

<table>
<thead>
<tr>
<th>Int. #</th>
<th>Int. ID</th>
<th>Ethn.</th>
<th>Gender</th>
<th>Age</th>
<th>Len of Acq</th>
<th>Geography</th>
<th>EH Pause Dur</th>
<th>EH Sp. Rate</th>
<th>Intr. Pause Dur</th>
<th>Intr. Speech Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CT</td>
<td>Euro. Am.</td>
<td>Male</td>
<td>30</td>
<td>0 days</td>
<td>Raleigh, NC</td>
<td>406 (N=132)</td>
<td>3.99 (N=150)</td>
<td>615 (N=35)</td>
<td>5.79 (N=77)</td>
</tr>
<tr>
<td>2</td>
<td>LM</td>
<td>Afr. Am.</td>
<td>Male</td>
<td>45</td>
<td>6 years</td>
<td>Raleigh, NC</td>
<td>455 (N=162)</td>
<td>4.01 (N=162)</td>
<td>395 (N=31)</td>
<td>4.15 (N=66)</td>
</tr>
<tr>
<td>3</td>
<td>CB</td>
<td>Afr. Am.</td>
<td>Female</td>
<td>62</td>
<td>&gt;15 years</td>
<td>NYC → Raleigh, NC 1976</td>
<td>368 (N=142)</td>
<td>3.99 (N=176)</td>
<td>494 (N=36)</td>
<td>3.75 (N=52)</td>
</tr>
<tr>
<td>4</td>
<td>TW</td>
<td>Euro. Am.</td>
<td>Female</td>
<td>29</td>
<td>8 years</td>
<td>Wilmington, NC</td>
<td>370 (N=156)</td>
<td>3.95 (N=204)</td>
<td>186 (N=28)</td>
<td>4.24 (N=59)</td>
</tr>
<tr>
<td>5</td>
<td>DC</td>
<td>Euro. Am.</td>
<td>Female</td>
<td>28</td>
<td>4 years</td>
<td>Illinois → Raleigh, NC 1986</td>
<td>364 (N=219)</td>
<td>4.35 (N=230)</td>
<td>260 (N=24)</td>
<td>4.52 (N=45)</td>
</tr>
</tbody>
</table>

Cullinan examined intra-speaker variation in pause durations, in addition to other speech features (such as rhythm and pitch differences between the interviews). She conducted this study primarily within the framework of SLAAP, and I am grateful to now be able to draw on her work to examine the stability and differences among EH’s pause durations and speech rates across these five interviews, and to further examine the effect of interviewer on pause and speech rate. The social and demographic characteristics of
Cullinan’s interviewers, along with EH and their pause and speech rate data are presented in Table 9.4.1.3

Figure 9.4.1 plots EH’s median pause duration and speech rate against those measures for her interviewers in each of the five interviews. We see here that despite differing values among – and differing genders and ethnicities of – her interviewers, EH’s speech rate and pause duration remain relatively constant. An ANOVA finds her speech rate differences significant ($p < 0.05$), but post-hoc t-tests show that she has a significantly faster speech rate in, and only in, her interview with DC. For the rest of her speech rate and pause data there are no significant differences. This stability on the part of EH contradicts the provocative findings of §9.3. EH provides evidence at least that a speaker’s speech rate and pause are not always influenced by her or his interactants.

![Pause Duration and Speech Rate in Five Interviews](image)

**Figure 9.4.1. Speech rate and pause duration for EH and her interviewers**

---

3 All pause durations and speech rates in Table 9.4.1 are median values.
9.5. A case study: Carissa is interviewing whom?

To examine this question further we turn to one final subset of the data. As I have mentioned in passing in earlier chapters, the recordings from Washington, DC come from an interesting source, with benefits for the project at hand (Kendall, Mallinson and Whitehead 2007; Mallinson 2007; Mallinson and Kendall forthcoming). They are sociological interviews with inner-city African American adolescents conducted for a Masters project by Carissa Froyum Roise, a white woman in her mid-twenties, originally from Minnesota, who was student in sociology at NC State University (Froyum Roise 2004).4

In 2001, Carissa worked as a counselor at a non-profit organization called “Urban Youth Network” (a pseudonym; henceforth UYN) in Washington DC. UYN is located in northeast DC and was founded in the 1970s to serve “at risk”, troubled, and homeless youth in DC. From summer 2001 until summer 2002, Carissa lived and worked at UYN. The next summer, in 2003, she returned to conduct an ethnographic study of the youths there. She observed 65 teenagers and interviewed 20 of them (9 boys, 11 girls). At the time the interviews were conducted, Carissa had worked at UYN for nearly two years, and had established herself as a trustworthy adult, counselor, and confidant to the youths (Froyum Roise 2004). The interviews were semi-structured; they were designed to elicit

---

4 It is a valid question to ask whether there are difficulties comparing the speech in these sociological interviews with the speech of sociolinguistic interviews. Since we do not have other data from Washington, DC with which to compare these interviews, we cannot conclusively say that the speech elicited in Carissa’s interviews is completely equivalent to the speech that would be elicited by sociolinguists. However, Mallinson and Kendall (forthcoming), Kendall, Mallinson and Whitehead (2007), and Mallinson (2007) have addressed this question and argued that these interviews in fact make for excellent sociolinguistic data.
data as to how the youth respond to the demands of inner-city life, and whether these responses differ by gender. The interviews were conducted in an office at the UYN center and were extremely similar to one another in terms of questions, scope, and setting.

Twelve of these twenty interviews – ten with girls and two with boys – are fully transcribed in SLAAP and have been included among the 104 speakers examined in Chapters 5 through 7. Since Carissa was the sole interviewer, had a similar relationship with all of the interviewees, and conducted relatively structured – and comparable – interviews, her speech in these interviews and the speech of her interviewees provide an excellent window further into variation at the interactional level. We begin by examining in closer detail the data for the interviewees before moving on to examine Carissa’s speech across these twelve interviews.

9.5.1. Pause and speech rate variation across twelve speakers

Table 9.5.1\(^6\) presents the median speech rate and median pause duration values for each of the twelve Washington, DC speakers for whom we have complete transcripts. Thus far, we have considered these twelve speakers as a “group”. They share regional affiliation and ethnicity. As Carissa describes (Froyum Roise 2004), they come from

\(^5\) As mentioned earlier, they comprise the entire set of Washington, DC speakers in the primary dataset. \(^6\) \((-\) \) and \((+)\) are used in these tables to indicate one standard deviation below or above the overall mean value, respectively.
similar socio-economic and educational backgrounds. They also share approximate age, all are between the ages of 12 and 17, and all but two are female.

### Table 9.5.1. Median pause duration and median speech rate for the twelve African American teenagers from Washington, DC

<table>
<thead>
<tr>
<th>Gender</th>
<th>Median Pause Dur. (ms.)</th>
<th>Median Speech Rate (σ/sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alayna Female</td>
<td>345 (-)</td>
<td>5.25 (+)</td>
</tr>
<tr>
<td>Asia Female</td>
<td>521</td>
<td>3.96</td>
</tr>
<tr>
<td>N = 1093</td>
<td>N = 1093</td>
<td></td>
</tr>
<tr>
<td>Calandra Female</td>
<td>477</td>
<td>4.21</td>
</tr>
<tr>
<td>Edwin Male</td>
<td>423</td>
<td>4.63</td>
</tr>
<tr>
<td>N = 299</td>
<td>N = 418</td>
<td></td>
</tr>
<tr>
<td>Elisa Female</td>
<td>445</td>
<td>4.87</td>
</tr>
<tr>
<td>Grace Female</td>
<td>410</td>
<td>4.89</td>
</tr>
<tr>
<td>Keisha Female</td>
<td>588 (+)</td>
<td>5.14 (+)</td>
</tr>
<tr>
<td>Latania Female</td>
<td>456</td>
<td>4.25</td>
</tr>
<tr>
<td>N = 1077</td>
<td>N = 1123</td>
<td></td>
</tr>
<tr>
<td>N = 848</td>
<td>N = 966</td>
<td></td>
</tr>
<tr>
<td>Shantell Female</td>
<td>526</td>
<td>4.44</td>
</tr>
<tr>
<td>Shawna Female</td>
<td>505</td>
<td>4.00</td>
</tr>
<tr>
<td>N = 1142</td>
<td>N = 1318</td>
<td></td>
</tr>
<tr>
<td>N = 441</td>
<td>N = 508</td>
<td></td>
</tr>
<tr>
<td>Shirlisa Female</td>
<td>416</td>
<td>4.29</td>
</tr>
<tr>
<td>Tad Male</td>
<td>552 (+)</td>
<td>3.24 (-)</td>
</tr>
<tr>
<td>N = 1449</td>
<td>N = 1939</td>
<td></td>
</tr>
<tr>
<td>N = 278</td>
<td>N = 401</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>472</td>
<td>4.43</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>69</td>
<td>0.58</td>
</tr>
<tr>
<td>μN = 791</td>
<td>μN = 919</td>
<td></td>
</tr>
<tr>
<td>σN = 391</td>
<td>σN = 446</td>
<td></td>
</tr>
</tbody>
</table>

Common sociolinguistic thinking might posit that these speakers, and the girls in particular, with shared ethnicity, and similar ages and socioeconomic backgrounds, form some sort of coherent speech group. Even if we don’t choose to simplistically consider these speakers as a *speech community* (cf. Patrick 2002), they likely fit the defining
characteristics of a *community of practice* (Eckert 2000). Regardless of what we call it, we would probably describe them as coming from relatively similar linguistic backgrounds. Yet, as we see from Table 9.5.1, there is a good deal of variation within the “group”.

![Figure 9.5.1. Distributions of speech rate and pause duration data](image)

Figure 9.5.1 displays the distributions of these speakers’ speech rate and pause medians in relation to the other 92 speakers examined so far. As in the earlier figures
(e.g. Figure 4.2.1) colors and shapes are used to differentiate speaker gender, ethnicity, and region. From this view, we see that the speech rate and pause data for the Washington, DC speakers are fairly spread out, but are contained within smaller ranges than the entire dataset. They are distributed on the slower half of the overall range of speech rates and they are distributed fairly centrally within the overall range of pause data.

Figure 9.5.2 plots the relationship between pause and speech rate for these speakers. As we see from this and the previous figure, there is a fairly large range of variation among the speakers’ pause durations and speech rates.

Figure 9.5.2. Plot of speech rate and pause duration for Washington, DC interviewees
The differences between speakers are confirmed as significant by an ANOVA ($p < 0.001$). We also find there is generally a linear inverse relationship between pause and speech rate among these speakers, especially if we ignore Keisha, who appears to be an outlier to this general pattern ($r^2 = 0.71$ without Keisha; though for all 12 speakers $r^2 = 0.20$). We did not find this inverse relationship in such a clear, linear pattern for the overall dataset, but this finding – of inverse linearity – is congruent with many of the results of early chapters (e.g., §7.2).

### 9.5.2. Pause and speech rate variation across twelve interviews

Table 9.5.2 presents median speech rate and pause duration measures for Carissa in each of the interviews with the Washington, DC speakers, along with the data for the speakers from Table 9.5.1. Here, we’ll focus on Carissa, the interviewer’s, data. Since all of the interviews were conducted in similar ways and Carissa had similar relationships with all of the adolescents, it seems commonsensical to approach these data with one of two expectations: Either Carissa’s speech (i.e., her pause and speech rate) will remain relatively constant between interviews or she’ll show evidence of accommodation to the individuals’ speech. As we noticed in §9.5.1, there is variation among the interviewees. In comparison to the interviewees, Carissa has a slightly shorter overall pause duration (413 ms compared to 472 ms) and slightly faster speech rate (4.93 $σ$/sec to 4.43 $σ$/sec).\(^7\)

---

\(^7\) These means are computed from the median values from each interview. They are very similar, however, to Carissa’s overall rates: Her overall median pause duration is 412 ms (calculated from the individual
Table 9.5.2. Median pause duration and median speech rate for interviewee and for Carissa, the interviewer, for each Washington, DC interview

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Interviewee Median Pause Dur. (ms.)</th>
<th>Interviewee Median Speech Rate (σ/sec.)</th>
<th>Carissa Median Pause Dur. (ms.)</th>
<th>Carissa Median Speech Rate (σ/sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alayna</td>
<td>345 (-)</td>
<td>5.25 (+)</td>
<td>363</td>
<td>5.22 (+)</td>
</tr>
<tr>
<td>Asia</td>
<td>521 N= 677</td>
<td>3.96</td>
<td>453</td>
<td>4.70</td>
</tr>
<tr>
<td>Calandra</td>
<td>477 N= 299</td>
<td>4.21</td>
<td>488 (+)</td>
<td>4.94</td>
</tr>
<tr>
<td>Edwin</td>
<td>423 N= 494</td>
<td>4.63</td>
<td>465</td>
<td>4.96</td>
</tr>
<tr>
<td>Elisa</td>
<td>445 N= 524</td>
<td>4.87</td>
<td>383</td>
<td>5.03</td>
</tr>
<tr>
<td>Grace</td>
<td>410 N= 1173</td>
<td>4.89</td>
<td>446</td>
<td>4.94</td>
</tr>
<tr>
<td>Keisha</td>
<td>588 (+) N= 1077</td>
<td>5.14 (+)</td>
<td>365</td>
<td>5.10</td>
</tr>
<tr>
<td>Latania</td>
<td>456 N= 848</td>
<td>4.25</td>
<td>388</td>
<td>4.99</td>
</tr>
<tr>
<td>Shantell</td>
<td>526 N= 1142</td>
<td>4.44</td>
<td>387</td>
<td>4.72 (+)</td>
</tr>
<tr>
<td>Shawna</td>
<td>505 N= 441</td>
<td>4.00</td>
<td>453</td>
<td>5.03</td>
</tr>
<tr>
<td>Shirlisa</td>
<td>416 N= 1449</td>
<td>4.29</td>
<td>313 (+)</td>
<td>4.53 (+)</td>
</tr>
<tr>
<td>Tad</td>
<td>552 (+) N= 278</td>
<td>3.24 (-)</td>
<td>455</td>
<td>4.94</td>
</tr>
<tr>
<td>Mean</td>
<td>µN= 791</td>
<td>µN= 919</td>
<td>µN= 330</td>
<td>µN= 484</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>σN= 391</td>
<td>σN= 446</td>
<td>σN= 81</td>
<td>σN= 125</td>
</tr>
</tbody>
</table>

Figure 9.5.3 attempts to show the relationships between Carissa’s median scores with those from her interviewees. Unlike the rates for EH in the previous section (cf. Figure 9.4.1), which were relatively stable across her interviews, Carissa manifests a range of variability. However, it is not obvious from the figure whether this variability

3,956 pause measurements), and her overall median speech rate is 4.92 σ/sec (calculated from the individual 5,802 speech rate measurements).
has a correlation with the each interviewee’s speech rate and/or pause realization. In a few cases, such as for Alayna, Carissa has near identical median values as her interviewees, but in many other cases there is no clear relationship, for example, as in her interview with Edwin.

![Pause Duration and Speech Rate in Twelve Interviews](image)

**Figure 9.5.3. Pause duration and speech rate comparison for Carissa and interviewees**

Figure 9.5.4 presents the speech rate and pause duration data presented earlier in Figure 9.5.1, but here I add the median measures for Carissa, the interviewer, for each of the twelve Washington, DC interviews. In the figure, the dotted lines connect Carissa’s median values with the values for her interviewees in each of the recordings. We note that for speech rate there is a correlation \( r^2 = 0.75 \) in the rank orders of Carissa’s speech rate measures with those of her interviewees, but not \( r^2 = 0.19 \) in the actual values of the measures.
That is, Carissa tends to speak with a higher speech rate when her interviewees do, but her overall rates are nonetheless contained within a 0.69 σ/sec range that is in the high-end of the range for her interviewees. She speaks at a faster speech rate than her slow interviewees and a slightly slower speech rate than her fastest talking interviewees. For
pause, we perhaps observe a similar tendency – Carissa often has pauses slightly shorter than her interviewees – but the actual correlation is quite low due to a number of outliers (rank order, \( r^2 = 0.12 \); actual values, \( r^2 = 0.05 \)).

A more complete analysis of the variation in these interviews than we have room for here, including an examination of morphosyntactic and phonological variation, would be necessary to fully understand the ways in which Carissa accommodates to her interviewees and designs her self in each of the interviews. While this is left for other (Mallinson and Kendall 2008) and future work, we take from this second case study the opposite sense gained from the first (in §9.4). Carissa does appear to modify her speech rate and pause realization in ways that relate at least somewhat to differences between her interviewees.8

9.6. In closing: Interactional variables vs. sociolinguistic variables?

Throughout this chapter we have seen evidence that a speaker’s overall (median) speech rate, and to a lesser extent pause duration, tend to vary in response to a number of aspects specific to their interactional contexts, including the gender of the interlocutors (§9.3.1) and the number of participants (§9.3.3), as well as accommodation and/or self-design oriented toward specific individuals (§9.5). We also found that all speakers may

8 As Mallinson and Kendall (2008) explore, some of these “modifications” may be found to correlate with speakers’ perceptions of their interlocutors more than they do with acoustic aspects of their interlocutors’ speech – as we have focused on here. Again, following up on this is left for work elsewhere.
not be influenced by their interlocutors (or that speakers may not be influenced by their interlocutors all the time), as appeared the case for EH (§9.4).

How do we ameliorate the contradicting findings of the case studies of EH (§9.4) and Carissa (§9.5)? Despite the fact that earlier in this work (Chapters 5-7) I have lumped interviewers and interviewees together as “subjects” of analysis, these two case studies highlight what is likely a major difference between these two interview roles. Interviewers have a stake in the interview process. As such, their willingness to and interest in accommodating to their interviewees is likely a major component of their active consciousness throughout the interview. Interviewees on the other hand, especially those like EH, who are participating as subjects as a “favor” to a friend, acquaintance, or even an unknown fieldworker from a distant university, may have little engagement in the interview process, and, as such, may make little effort to accommodate to their interviewer(s) (cf. Hazen 2000). Cullinan (2007: 6, fn. 8) describes that EH at one point reported that the interviews “wore her out”. It is, perhaps, not surprising then that her speech rates and pauses were very little impacted by her interviewers.

From the findings of this chapter it seems likely that, at least in some cases, these interactional effects appear to trump the other predictors. For example, the fixed-effect model presented in Table 9.3.2 illustrates that despite the earlier findings (e.g., in Table 7.2.1) that a speaker’s regional background, gender, and ethnicity influence their speech rate – and that there is a predictive inverse relationship between pause duration and speech rate – speech rate may, in fact, principally be an effect of the interviewer’s gender.
As I declared above, if this sort of observation holds in future work, it will have far reaching implications for our understanding of and ways for studying linguistic variation.

The discussions in this chapter, while focused on speech rate and pause, raise some questions that are important for the larger sociolinguistic picture. To what extent do these findings for pause and speech rate inform our understanding of “normal” sociolinguistic variables? Further, to what degree are pause and speech rate really different from “normal” sociolinguistic variables? When examined from the perspectives taken here, “normal” variables are often found to be constrained in similar ways, with realizations that are at least somewhat dependent on interactional aspects of the speech event in which they are produced beyond social attributes of the speaker and monolithic conceptions of style (Rickford and McNair-Knox 1994; Hazen 2000; Schilling-Estes 2004a; Mallinson and Kendall 2008).

We turn now to the final chapter of Part 2 to shed some additional light on these questions.
10. The relationship(s) between speech rate and pause and “traditional” sociolinguistic variables

10.1. Introduction

Thus far, I have examined speech rate and pause as sociolinguistic variables in their own right, correlating with social characteristics of speakers (Chapters 5-7), and also with discourse contextual aspects of their matrix conversations in terms of stylistic and register variation (Chapter 8) and interactional effects (Chapter 9). We have also seen in the examination of register variation that changes in pause and speech rate do not appear to correlate with overall changes in speakers’ productions of phonological and morphosyntactic variables categorically in a single direction (Chapter 8). In this chapter, I turn to an area where pause and speech rate have been traditionally considered within sociolinguistic work, dating back to Labov’s foundational (1966) study of English in New York City: I consider these speech features as paralinguistic cues and as indicators of attention to speech (Labov 1972). I attempt to expand the most often impressionistically implemented and somewhat theoretically shallow concept of paralinguistic cue by drawing on the long line of research in psycholinguistic traditions that connects these temporal sequencing features with cognition and language production. First though, I start by framing this project in terms of a broader pursuit for
sociolinguistics – an attempt to put language variation studies in dialogue with the investigation of language and cognition.

10.2. Performance and production

Much recent sociolinguistic work has focused on performance (e.g., Schilling-Estes 1998) and dialect stylization (e.g., Coupland 2001, 2007; Podesva 2007). Recent work has also emphasized the role that salient variables play in persons’ (local) identity management without necessarily putting this in terms of performance (e.g., Pappas 2008; Van Herk, Childs, and Assiri 2008; Snell 2008). In other words, a common finding has been that younger speakers in post-isolated communities use the traditional local forms less than their elder counterparts, but, often times in these situations, the younger speakers do make at least symbolic use of those local forms, with speakers invoking or performing salient local dialect features in ways that assert localized identities (e.g., Wolfram and Schilling-Estes 1995; Van Herk, Childs, and Thorburn forthcoming; cf. Schilling-Estes 1998).

An area that has been under-addressed in the literature is a deeper questioning of just how we might be sure (or better “prove”) that the use or suppression of particular linguistic features are performed. That is, what exactly – in a cognitive sense – is performance? It seems to me that interpreting a particular stretch of speech as performance has deeper (and more interesting) implications than sociolinguists often pursue. Namely, by considering the use of lower-frequency, “traditional” forms as
performance (or the non-use of higher-frequency tradition forms as some sort of suppression) we are implicitly determining their use (or non-use) to be the result of conscious or semi-conscious control. A corollary of this sort of interpretation is that performance is a non-native or semi-native linguistic strategy – that is, that the local features are less integrated into these speakers’ native grammars than they are in the grammars of the older, local-dialect speakers. If so, might there be empirical evidence for the performed nature of these less-used variable productions?

Since Labov’s original explorations of the correlation of “paralinguistic cues” with attention to speech style shifting (Labov 1966, 1972), the complex relationship between linguistic variables and speech features such as pause and speech rate has rarely been examined in a systematic, quantitative way. The main exception to this seems to be in studies of syllable-coda consonant cluster reduction (CCR), also called final stop or t/d deletion, where numerous studies (e.g., Guy 1980) have found following pause to be an important environment in the patterning of consonant cluster reduction. In other words, analysts such as Guy have coded for when the following environment is “quiet” (code Q). In some cases, such as Fasold’s (1972) study of African American English in Washington DC and Wolfram, Childs, and Torbert’s (2000) study of CCR in African American English in Hyde County, pause appears to pattern along with consonants for this variable, but in others (e.g., the Cherokee Sound Anglos in Wolfram, Childs and Torbert’s 2000 analysis) pause patterns more similarly to vowels. In Guy’s own (1980) analysis, pause emerges as a factor that differentiates the white dialects in New York City versus Philadelphia. Interestingly, Guy (1980) notes that pause is outside of the sonority
hierarchy. Therefore, it is not linguistically constrained and we might consider differences in variable rates depending on pause to be purely socialized.

In this usage however, pause is treated as a structural category – it is the absence of a segment in the following position. There have yet to be examinations that look at the relationship between pause variability and variable realization – that is, CCR and t/d deletion studies have not considered the importance of pause length. Guy (1980) has, however, voiced an interest in the rate of speech and said that the “probability of deletion apparently increases in proportion to the rate of speech” (1980: 9). He noted that at the time there was not yet a reliable way to measure rate or speech, so he did not include it in his analysis.  It is this sort of more “paralinguistic” assessment of both pause and speech rate that I want to pursue further here.

At the same time, recall (from §5.2) that pause has been found to correlate with cognitive processing (Goldman-Eisler 1968). Understanding attention to speech and pausing as not only related to one another but also related to more general cognitive and language production processes may lead to a richer conception of linguistic variation. It also reminds us that there may be methods arising from the productive work on speech timing within the psycholinguistic literature that could be productively incorporated into sociolinguistic analysis.

1 Guy’s recent work (personal communication) has returned to this question, and found some evidence in support of the relationship between rate of speech and deletion.
10.3. The “Henderson graph” redux: A metric for the analysis of paralinguistic cues to variable realization

As he himself has noted (personal communication), Guy’s (1980) problems with accurately measuring rate of speech are no longer the case. Through instrumental techniques, the mechanics of measuring rate of speech and pause are relatively straightforward (as discussed in §6.4). Nonetheless, asking how speech rate and pause may interact with, effect, or predict variable production is not at all a clear issue. For example, in attempting to map the realization of pauses against the realization of a particular variant form, should one measure the duration of the most recent pause, or the duration since the most recent pause, or both? In searching for these sorts of possible correlations how do we know when or if we have discovered the most meaningful relationship? How do we know when or if we have found merely a symptom of the linguistically or cognitively meaningful relationship instead of the relationship itself? To investigate this, I adopt a broader-based metric than simply pause or speech rate, and revisit the Henderson graph, first mentioned in §3.4.

As discussed earlier, the Henderson graph (Figure 3.4.4, reproduced here as Figure 10.4.1) is a representation of a speech event in which talk-time is plotted on the x-axis while pause-time extends along the y-axis. Changes in the characterization of the talk are viewable as changes in the slopes of sections of the talk. The slope measurements themselves provide a quantitative metric that can be correlated with the speakers’ variable productions.
Additionally, the Henderson segments – the stretches of talk by a speaker that are fitted to a single slope line – provide an empirical, etic segmentation of the discourse that allow us to test other speech timing features (like pause duration, speech rate, and even length of “segment”) against variable realizations. Without the Henderson graph slope lines to segment the discourse, we would have to determine other means to assess matrix characteristics like speech rate. That is, other than looking at the single phonetic utterance (i.e. the SLAAP transcript line) that a variable is realized within, the only other options appear to be fairly arbitrary – such as selecting an arbitrary number of words or length of time on either side of the variable production. In many ways, the approach to

**Figure 10.3.1. Example of a Henderson graph (Levelt 1989: 127)**
segmenting the talk into units using the changes in slope within the Henderson graph appears more organic.²

Table 10.3.1. Temporal sequencing variables

<table>
<thead>
<tr>
<th><strong>SLOPE</strong></th>
<th>Best-fit slope over the current Henderson segment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ΔSLOPE</strong></td>
<td>Change in slope from previous Henderson segment</td>
</tr>
<tr>
<td><strong>SLOPECOMP</strong></td>
<td>Tertiary variable based on the comparison of a given slope and that speaker’s mean Slope (low, norm, high)³</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>Correlation coefficient for line</td>
</tr>
<tr>
<td><strong>DUR</strong></td>
<td>Duration of Henderson segment</td>
</tr>
<tr>
<td><strong>PAUSEN</strong></td>
<td>Number of pauses within segment</td>
</tr>
<tr>
<td><strong>PAUSEDUR</strong></td>
<td>Median pause duration within segment</td>
</tr>
<tr>
<td><strong>ARTRATE</strong></td>
<td>Median articulation rate within segment (σ/second, not including pauses)</td>
</tr>
<tr>
<td><strong>SpkRate</strong></td>
<td>Overall speaking rate for segment (total # σ / total duration of segment)</td>
</tr>
</tbody>
</table>

² Admittedly, this is an empirical question and I have not pursued it in as much depth as I would like. For speech rate, in particular, it would be fairly straightforward to extract speech rate for each variable realization’s matrix phonetic utterance (i.e. transcript line) from SLAAP and then compare the effect of that measure against the predictive success of the Henderson segment based measures (see Table 10.4.1). For some variables, such as CCR, it seems reasonable to think that this more “local” measure of speech rate might have some success. To jump ahead to the case study analysis presented in §10.4, however, readers will note that speech rate-based measures do not surface as significant predictors for the variables examined (with one exception explained in footnote 15 below), so pursuing this question here did not seem warranted. To make one final comment on this digression: At the language production level, it seems to me that the difference between “locally” based measures of speech rate and the broader Henderson graph based sequential temporal measures is parallel to predicting an articulatory effect versus a cognitive/processing effect on variable realization. That is, if we expect that CCR is the outcome of or affected by articulatory processes, than we might want to pursue the utterance-level speech rate value as an independent variable. If we expect that larger language production processes are at work, I would argue that the Henderson graph based metrics are more appropriate on a theoretical basis.

³ With the exception of SLOPECOMP all of the temporal sequencing measures are continuous variables.
Table 10.4.1 lists the temporal sequencing variables I am considering based on the Henderson graph metric. Other measures could surely be generated from the method, but we will limit our inquiry to those in Table 10.4.1. Note that the SLOPE, ΔSLOPE, SLOPECOMP, and R² variables are directly related to the slope of the best-fit line for a Henderson segment. The other five variables, DUR, PAUSEN, PAUSEDUR, ARTRATE, SpkRATE,⁴ are determined based on the stretch of time within the Henderson segment; they are not measures derived from the Henderson graph itself. These variables (e.g., ARTRATE, the median articulation rate within a stretch of talk) are used to determine if features like rate of speech correlate with variable productions. While not derived directly from the Henderson graph, they are dependent on the segmentation of the talk by its changes in slope.

The first three variables of Table 10.4.1, SLOPE, ΔSLOPE, SLOPECOMP, are in bold-face because, in addition to relating directly to the slope lines of the Henderson graph, they allow us to ask, in a way, certain questions about the relationship between variable realizations and the sequential temporal aspects of their matrix talk. Namely:

- **SLOPE**: Since steeper slopes are assumed to indicate more hesitant speech (Henderson et al. 1966) – what we might consider as possibly more self-conscious talk – do steeper slopes indicate higher likelihood of salient linguistic forms?
- **ΔSLOPE**: Can changes in slope be understood as paralinguistic cues to changes in speech style?

---

⁴ Recall the distinction between articulation rate (ARTRATE) and speaking rate (SpkRATE) from §6.2.
• **SlopeComp**: Does the distance of a particular slope (over a stretch of talk) from a speaker’s mean slope predict less typical variable productions in that talk?

### 10.4. A case study: Language and identity in Petty Harbour, Newfoundland

In order to test this metric – and more broadly the hypothesis that we can find correlations between processing or “consciousness” (Chafe 1994) and the realizations of sociolinguistic variables (and that these correlations further tell us about the salience or intentionality of speakers’ use of marked linguistic forms) – I turn now to a brief analysis of data from Petty Harbour, a post-isolated, urbanizing community 15 km from St. John’s, the capital of Newfoundland.5

Van Herk, Childs, and Thorburn (forthcoming) examine language change and identity in Petty Harbour and present data on two variables, the phonological variable of interdental fricative stopping (dis* ting for this thing) and the morphosyntactic variable of non-standard verbal –s marking (e.g., I like* having fun with the kids). Examining 24 sociolinguistic interviews with Petty Harbour natives, they find that the younger speakers, in particular the young women, use the local features the least, while the older speakers, especially the men, use the local features the most. For these variables, they explain their data as indicating that “the nonstandard variant acts as a salient (though

---

5 I am extremely grateful to Gerard Van Herk, Becky Childs, and Jennifer Thorburn for sharing some of their data with me and for discussing this project with me. While I have not included these data in the macro-level analyses of the earlier chapters, thanks to Van Herk, Childs, and Thorburn, I was able to examine six of the Petty Harbour interviews within SLAAP.
perhaps not fully consciously deployed) marker of traditional Newfoundland identity” (Van Herk et al. forthcoming).

Van Herk et al. understand this variation in terms of saliency and, classifying voiced interdentals into two groups (with variable distributions as shown in Table 10.4.1), high-frequency +Function words and lower frequency -Function words, they propose that the greater rate of stopped voiced interdentals for high-frequency (+Function) words like the and this is due to their being “less salient and thus less suppressed” (forthcoming).

| Petty Harbour non-standard (ð) use (Van Herk et al. forthcoming: Table 2) |
|-----------------|-----------------|-----------------|
| Older women (60+) | 80.3 % | 13.9 % |
| Older men (60+) | 88.3 % | 95.6 % |
| Middle women (30-60) | 38.8 % | 15.1 % |
| Middle men (30-60) | 74.5 % | 69.1 % |
| Younger women (under 30) | 42.8 % | 8.9 % |
| Younger men (under 30) | 38.5 % | 29.4 % |
| Total | 60.5 % | 38.7 % |

But, we might also alternatively suppose the contrary – that high-frequency words like the and this are more salient by virtue of their frequency. That is, perhaps their high occurrence makes them more available for identity work. In this view, we might suppose that their higher rates of (ð) stopping are performative acts, intentionally (or not) indexing
a more localized identity. In sum, based purely on the rates of use of the variables, alternative and contradictory conclusions can be drawn. So, how can we pick between one of these two alternative understandings?6

To investigate this problem, I here examine a subset of Van Herk et al.’s interviews and examine two variables: The stopping of (voiced) interdentals (ð), e.g., dis for this, and velar nasal fronting (ing), e.g., talkin’ for talking. I had hoped to also examine non-standard verbal –s marking, but the feature proved to be too rare to be systematically studied in the subset of data I am examining.7 Following from Van Herk et al.’s work, (ð) here is considered to be a salient variable. (ing) was not examined by Van Herk et al. and I use it as a control variable, assuming that its use does not index any sort of local identity. As is often found to be the case (going back at least to Fischer 1958), it is assumed however that (ing) varies with formality, with more full –ing realizations occurring in more formal speech and more –in’ occurring in less formal speech (also see Labov 1966, 2001). Table 10.4.2 provides a summary of the seven speakers examined here. Three are old males, born between 1938 and 1944, and four are young females, born between 1984 and 1988. I have limited this analysis to these two demographic subsets because according to Van Herk et al.’s work they pattern on opposite extremes. The “Included” row in the table indicates the length of each interview’s segments that were transcribed and included in this analysis.

---

6 Van Herk (personal communication) cites historical and comparative evidence that high-frequency function words, like the and this, maintain traditional pronunciations longer than less frequent, content, words. It is clear that there are broader sources of evidence that bear on answering this question, but I would like to focus here on the empirical, local evidence.

7 In the subset of data examined for this case study, I obtain only 1/15 instances of non-standard –s attachment for old males, and 4/36 for young females. Van Herk et al. obtained only 61 –s marked verbs (5.6% of the 1,090 potential cases) in their entire sample.
Table 10.4.2. Subset of data examined here from Van Herk et al. (forthcoming)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Young Females</th>
<th>Old Males</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d f s z</td>
<td>h H K</td>
</tr>
<tr>
<td>YOB</td>
<td>1984 1987</td>
<td>1944 1941</td>
</tr>
<tr>
<td>Included</td>
<td>10 min 30 min</td>
<td>20 min 10 min 7.5 min</td>
</tr>
</tbody>
</table>

10.4.1. Methods for temporal analysis

After transcribing the segments and incorporating them into SLAAP (see Chapter 3), I extracted and coded the transcribed portions of the interviews for (ð) and (ing) using SLAAP’s variable tabulation feature to record accurate timestamps along with the variable codes. Since I am interested in situating the variable productions within their matrix speech, I have not limited the number of tokens by type. Instead, I have coded all (ing) variables that occur within the 1.6 hours of transcribed speech for the Petty Harbour speakers. For (ð), which is extremely common in normal English speech, I limited the variable context to the following five lexical items, all members of Van Herk et al.’s +Function class: the, this, that, these, and those. For these words, I coded all instances of (ð) that were either perceptually fricated or stopped. I did not count instances that I perceived as affricated, assibilated, zero, or otherwise assimilated to the previous

---

8 Speaker s, a Petty Harbour native, was the interviewer for speaker z. Both are included as speakers here. A combined 19.2 minutes was examined from the recording of their interview.
consonant (e.g., on ne bus). I also did not count any instances that were in potentially ambiguous environments (such as at the, with the), regardless of whether the realization was clear or ambiguous.

Using SLAAP’s Henderson graph feature, I generated Henderson graphs and best-fit slope lines for all of the transcribed portions of the Petty Harbour interviews. This is illustrated by the screenshot from SLAAP in Figure 10.4.1.

The slope lines are generated by SLAAP based on points I select on the graph. The generation of the slopes is a manual and time-consuming process, but SLAAP stores the slope values and other metrics in a spreadsheet as the slopes are calculated. When this
task was completed for each speaker, I downloaded the generated spreadsheet obtaining the temporal sequencing variables outlined in Table 10.3.1, above.

By virtue of the fact that both sets of variables – linguistic and temporal sequencing – are time stamped in SLAAP, the sociolinguistic variable data for (ð) and (ing) were merged with the temporal variables using scripts that I wrote in R (R Development Team 2008).

10.4.2. (ð) analysis

The overall data for (ð) are presented in Table 10.4.3. Van Herk et al.’s (forthcoming) data for their +Function class of high-frequency words are also included in the right-hand column for comparison.

Table 10.4.3. % stopped and Ns for (ð)

<table>
<thead>
<tr>
<th></th>
<th>the</th>
<th>that</th>
<th>this</th>
<th>those</th>
<th>these</th>
<th>totals</th>
<th>Van Herk et al. +Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Males</td>
<td>80/155</td>
<td>25/55</td>
<td>5/7</td>
<td>2/3</td>
<td>0</td>
<td>112/220</td>
<td>N = 163 80.3%</td>
</tr>
<tr>
<td></td>
<td>51.6%</td>
<td>45.4%</td>
<td>71.4%</td>
<td>66.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young Females</td>
<td>61/155</td>
<td>16/76</td>
<td>6/13</td>
<td>0/1</td>
<td>0/2</td>
<td>83/247</td>
<td>N = 175 42.8%</td>
</tr>
<tr>
<td></td>
<td>39.4%</td>
<td>21.1%</td>
<td>46.1%</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>141/310</td>
<td>51/131</td>
<td>11/20</td>
<td>2/4</td>
<td>0/2</td>
<td>190/467</td>
<td>N = 946 60.5% (all spkrs)</td>
</tr>
<tr>
<td></td>
<td>45.5%</td>
<td>33.1%</td>
<td>55.0%</td>
<td>50.0%</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9 The total rate of stopped (ð) from Van Herk et al.’s data (60.5%) is from all tokens for all speakers from six demographic categories. The rate of stopped (ð) for just the two groups relevant here is 63.5%, N = 321.
Various regression models were fitted to the data in an attempt to gain a sense of the best predictors for (ð). The best model, from a step-up/step-down analysis using Rbrul (Johnson 2008, 2009), is presented in Table 10.4.4. As expected from Van Herk et al.’s (forthcoming) findings, the demographic category of “old male” versus “young female” is found to be significant, with old males favoring and young females disfavoring stopped (ð).

<table>
<thead>
<tr>
<th></th>
<th>p = 0.0001</th>
<th></th>
<th>Stop/Stop+Fric</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE+GENDER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old Male</td>
<td>0.348</td>
<td>0.59</td>
<td>220</td>
</tr>
<tr>
<td>Young Female</td>
<td>-0.348</td>
<td>0.42</td>
<td>245</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WORD</th>
<th>p = 0.0346</th>
<th></th>
<th>Stop/Stop+Fric</th>
</tr>
</thead>
<tbody>
<tr>
<td>this</td>
<td>0.518</td>
<td>0.65</td>
<td>20</td>
</tr>
<tr>
<td>those</td>
<td>0.024</td>
<td>0.54</td>
<td>4</td>
</tr>
<tr>
<td>the</td>
<td>0.014</td>
<td>0.53</td>
<td>310</td>
</tr>
<tr>
<td>that</td>
<td>-0.348</td>
<td>0.40</td>
<td>131</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>25</td>
<td>(these excluded as a knockout)</td>
</tr>
</tbody>
</table>

Not selected as significant: SLOPE, ΔSLOPE, SLOPECOMP, R², DUR, PAUSEN, PAUSEDUR, ARTRATE, SPKRATE

As mentioned above, Van Herk et al. found that the status of a word as a high-frequency, +Function, word as a opposed to a -Function word had a major impact on (ð) realization,

---

¹⁰ For fixed effects, Rbrul can report both log-odds values and GoldVarb-like factor weights (here reported as “uncentered weights”). Unlike the regression models reported in earlier chapters (which were for continuous response variables), I report both here. Results for continuous predictors (like SLOPE) or response variables (like speech rate and pause in earlier chapters) cannot be reported or understood in terms of factor weights so are only given as log-odds values.
with the +Function words favoring (ð) stopping, especially for the females (cf. Van Herk et al. forthcoming, Tables 2 & 3). Here we are only examining words in the +Function category, but we note the model indicates that the actual word type is significant ($p < 0.05$). This most strongly favors stopped (ð), those and the slightly favor stopped (ð), and that disfavors stopped (ð). These was excluded from the analysis as a knockout, with only two tokens (neither stopped) for the young females and none for the old males. Interestingly here, we also note that it is the less frequent lexical items, this and those, that favor stopping most strongly. None of the sequential temporal variables are found to be significant predictors in this model.

“Grammatical” language differences are best found by comparing models for each of the groups of interest (Tagliamonte 2006). So, at this point we ask: What happens if we separate the two demographic groups? Do they show the same pattern? Various regression models for the old male speakers fail to yield any factor groups or factors of significance when word, SLOPE, ΔSLOPE, SLOPECOMP, $R^2$, DUR, PAUSEN, ARTRATE, and SPKRATE are run as potential predictors. This lack of significance is actually quite useful information – for the old males we do not even find the lexical pattern found when the data were not separated by demographic group, with certain items within Van Herk et al.’s +Function category favoring stopping to differing degrees. That is, for the old males, it appears that stopping is equally likely regardless of (+Function) word-frame and regardless of the temporal sequencing character of the matrix talk.

For the young women alone, however, a regression model does obtain significance. This best model is shown in Table 10.4.5. As we found for all the speakers
combined (in Table 10.4.4), word is significant – and even more so here than in the overall model ($p < 0.01$) – with the same basic pattern, although here *those* is excluded from analysis in addition to *these* as it is a knockout when only the young females are considered. More importantly for the present investigation, $\Delta\text{SLOPE}$ is found to be significant as well ($p < 0.05$). This model indicates that positive changes in slope (i.e. relative increases in hesitancy) lead to increased probabilities of ($\delta$) stopping. It must be noted that, although the 0.822 log-odds value associated with a +1 change in slope appears quite large, the mean absolute value of the $\Delta\text{SLOPE}$ scores is only 0.26 for all speakers so the strength of this effect may not be quite as large as it seems from first glance at the table.\footnote{A t-test comparing the $\Delta\text{SLOPE}$ values for fricated ($\delta$) (mean = -0.11) versus stopped ($\delta$) (mean = -0.01) does not yield significance, but obtains the almost significant value of $p = 0.066$.}

### Table 10.4.5. Regression analysis for ($\delta$) for young females

<table>
<thead>
<tr>
<th>WORD</th>
<th>$p = 0.0051$</th>
<th>Log-odds</th>
<th>Uncentered Weight</th>
<th>Tokens</th>
<th>Stop/Stop+Fric</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>this</em></td>
<td>0.550</td>
<td>0.65</td>
<td>13</td>
<td>0.462</td>
<td></td>
</tr>
<tr>
<td><em>the</em></td>
<td>0.219</td>
<td>0.57</td>
<td>155</td>
<td>0.394</td>
<td></td>
</tr>
<tr>
<td><em>that</em></td>
<td>-0.769</td>
<td>0.33</td>
<td>76</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>32</td>
<td>(these &amp; those excluded as knockouts)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| $\Delta\text{SLOPE}$ | $p = 0.0261$ | Log-odds +1 0.822 | (continuous predictor) |

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total $N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>299.017</td>
<td>4</td>
<td>-0.581</td>
<td>0.340</td>
<td>244</td>
<td></td>
</tr>
</tbody>
</table>

Not selected as significant: SLOPE, SLOPECOMP, $R^2$, DUR, PAUSEN, ART RATE, SPKRATE
Nonetheless, this indicates that a stretch of talk by a young female speaker with steeper slope – that can be deemed as “more hesitant” – than the stretch of talk preceding it is more favorable for stopped (ð) than a stretch of talk with a less steep slope than the previous segment. This relationship is illustrated in Figure 10.4.2. (a) in the figure provides an example of a situation with a ΔSLOPE of 2.5, which would have a log-odds likelihood of 2.055 of (ð) realized as a stop, [d]. (c), on the other hand, with a negative ΔSLOPE, of -0.7, would have a greater likelihood of fricated (ð), [ð], since its log-odds value would be -0.575. Finally, (b), in between the other two examples with its slightly positive ΔSLOPE of 0.5 would have a log-odds likelihood of stopped (ð), [d], of 0.411.

![Figure 10.4.2. Favorability of stopped (ð) according to ΔSLOPE](image)

10.4.3. (ing) analysis
(ing) was selected for analysis because I assumed it would make a good control variable. Unfortunately, as we see from Table 10.4.6, the old males used –in’ almost categorically, a fact which mitigates the utility of the comparison.\textsuperscript{12} Altogether the old males have only 4 full –ing realizations out of 88 total tokens, all nouns and pronouns (everything x2, nothing x1, and pudding x1).\textsuperscript{13} Nonetheless, we still examine (ing) here hoping that it may provide us a window into the relationship of sequential temporal features of speech and sociolinguistic variables.

\begin{table}[h]
\centering
\caption{\% fronted and Ns for (ing)}
\begin{tabular}{|c|c|}
\hline
          & –in’ \tabularnewline
\hline
Old Males & 84/88 \tabularnewline & 95.5\% \tabularnewline \hline
Young Females & 86/176 \tabularnewline & 48.9\% \tabularnewline \hline
Totals   & 170/264 \tabularnewline & 64.4\% \tabularnewline \hline
\end{tabular}
\end{table}

Table 10.4.7 presents the best regression model from a step-up/step-down Rbrul run for all of the (ing) data. Demographic category is found to be highly significant, as we would expect due to the near categorical nature of the old males’ –in’ use. Grammatical category is also found to be highly significant. This is congruous with

\textsuperscript{12} Categoricity, of course, makes regression analysis pointless, since all independent variables will equally likely favor the categorical outcome.
\textsuperscript{13} And, we note that everything is realized with full –ing most of the time for most speakers. In this dataset, however, one of the young females, f, does notably have two fronted tokens of everything.
previous findings on (ing), such as Hazen’s recent (2008) study of (ing) in West Virginia, as is the general pattern of favorability for the different grammatical categories.

Progressives and gerund-participles (e.g. …without getting caught) favor –in’, while other grammatical categories favor –ing.\textsuperscript{14}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{AGE+GENDER} & \textbf{p = 3.68x10\textsuperscript{-15}} & \textbf{Log-odds} & \textbf{Uncentered Weight} & \textbf{Tokens} & \textbf{-in‘-in’+ -ing} \\
\hline
Old Male & & 1.671 & 0.90 & 88 & 0.941 \\
Young Female & & -1.671 & 0.25 & 176 & 0.471 \\
Range & & & 65 & & \\
\hline
\textbf{GRAMMATICAL CAT} & \textbf{p = 1.63x10\textsuperscript{-6}} & \textbf{Log-odds} & \textbf{Uncentered Weight} & \textbf{Tokens} & \textbf{-in‘-in’+ -ing} \\
\hline
Progressive & & 1.284 & 0.72 & 111 & 0.820 \\
Gerund-Part & & 0.487 & 0.54 & 20 & 0.850 \\
Gerund & & -0.159 & 0.38 & 48 & 0.542 \\
Noun & & -0.607 & 0.28 & 72 & 0.458 \\
Adjective & & -1.004 & 0.21 & 13 & 0.231 \\
Range & & & 51 & & \\
\hline
\textbf{SLOPE} & \textbf{p = 0.0123} & \textbf{Log-odds} & \textbf{(continuous predictor)} & \textbf{+1} & \textbf{-1.800} \\
\hline
\textbf{Model} & & & & & \\
Deviance & 237.068 & df & 7 & Intercept & 1.816 \\
 & & & Mean & 0.644 & Total N \\
 & & & & & 264 \\
\hline
\end{tabular}
\caption{Regression analysis for (ing) for all speakers}
\end{table}

Not selected as significant: PRECEDING ENV, FOLLOWING ENV, ΔSLOPE, SLOPECOMP

Not submitted for this run:\textsuperscript{15} R\textsuperscript{2}, DUR, PAUSEN, PAUSEDUR, ARTRATE, SPKRATE

\textsuperscript{14} For comparison, Hazen (2008) reports the hierarchy (factor weights in parentheses): Progressive (.72) > Gerund-Participle (.65) > Noun (.25) > Adjective (.18) > Gerund (.15). The general pattern of progressives and gerund-participles favoring –in’ (> .50) and gerunds, nouns, and adjectives disfavoring (< .50) is found here in the Petty Harbour data, but note that the actual orders of the disfavoring factors differs in the models for the Petty Harbour speakers (Tables 10.4.7 & 10.4.8) from Hazen’s West Virginia data.

\textsuperscript{15} For the regression models presented in Tables 10.4.7 and 10.4.8, I did not include a number of the sequential temporal variables. I did this because a number of those variables (such as PAUSEN and SPKRATE) actually do obtain significance in the model. However, those models which include the fuller set of potential predictors are “messier” – they have higher deviance or higher degrees of freedom and, overall, seem to be less good models. They also select sequential temporal variables that we can understand as contained within the SLOPE measure. For example, SPKRATE as a rate of speech measure includes the durations of pauses and PAUSEN is simply the number of pauses within the Henderson segment. Both of
Most importantly, SLOPE is found to be significant ($p = 0.01$). Increases in the Henderson slope lead to decreases in the log-odds of \( -i\text{n} \) realization. In other words, for speech that is more hesitant, the likelihood of a fully velar, \( -i\text{ng} \), realization is higher. This is not at all surprising (cf. the early findings of Fischer 1958; also Labov 1966, 2001) and gives support to the success of the Henderson slope method as a quantitative metric for attention to speech.

As mentioned above, we cannot readily compare differences in the significant factors for the two demographic groups, since the old males are near categorical in the fronting of velar nasals. It is not meaningful to run logistic regression on the old males’ group on their own, since – as a glance at the data indicates – only four (4.5%) of the (ing) tokens are \( -i\text{ng} \) for the old males. Running a logistic regression on just the young females, however, is possible, and doing so yields the model presented in Table 10.4.8. The effects of grammatical category and SLOPE are still found to be significant, and we find here that in fact SLOPE is more significant ($p = 0.005$) and has a stronger prohibitive effect on fronting than in the model for all speakers.

A t-test comparing the Henderson slope values for the young females’ \( -i\text{n} \) realizations (mean slope = 0.21) versus their \( -i\text{ng} \) realizations (mean slope = 0.30) yields significance ($p = 0.01$), further confirming the fact that there are significant differences in the ways that the slopes distribute with respect to the (ing) realizations. In sum, higher slopes correlate with full velar nasal productions for the young females, even more

\[\text{these measures seem to me robustly captured by the SLOPE metric. Using a “messier” model that uses the two (interacting) predictors instead of one is less preferable.}\]
strongly than they did for both groups together in the combined model. The strength of this effect appears quite large.\footnote{Note that a slope of only about 0.6 (log-odds = -1.33) is enough to counteract the favoring fronting effect of progressive grammatical forms (log-odds = 1.34). We must also note, however, that only about 7\% of the young females’ slopes for (ing) are actually above 0.6, so clearly this is not to say that grammatical category is often overridden as a significant factor in (ing) realization.}

**Table 10.4.8. Regression analysis for (ing) for young females**

<table>
<thead>
<tr>
<th>GRAMMATICAL CAT</th>
<th>p = 1.86x10^{-5}</th>
<th>Log-odds</th>
<th>Uncentered Weight</th>
<th>Tokens</th>
<th>-in'/-in' '+'-ing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressive</td>
<td></td>
<td>1.341</td>
<td>0.73</td>
<td>71</td>
<td>0.718</td>
</tr>
<tr>
<td>Gerund-Part</td>
<td></td>
<td>0.127</td>
<td>0.45</td>
<td>6</td>
<td>0.500</td>
</tr>
<tr>
<td>Gerund</td>
<td></td>
<td>-0.199</td>
<td>0.37</td>
<td>34</td>
<td>0.353</td>
</tr>
<tr>
<td>Noun</td>
<td></td>
<td>-0.328</td>
<td>0.34</td>
<td>52</td>
<td>0.327</td>
</tr>
<tr>
<td>Adjective</td>
<td></td>
<td>-0.942</td>
<td>0.22</td>
<td>13</td>
<td>0.231</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SLOPE</th>
<th>p = 0.0053</th>
<th>(continuous predictor)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-odds</td>
<td></td>
</tr>
<tr>
<td>+1</td>
<td>-2.216</td>
<td></td>
</tr>
</tbody>
</table>

Model

<table>
<thead>
<tr>
<th>Deviance</th>
<th>df</th>
<th>Intercept</th>
<th>Mean</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>208.887</td>
<td>6</td>
<td>0.155</td>
<td>0.489</td>
<td>176</td>
</tr>
</tbody>
</table>

Not selected as significant: PRECEDING ENV, FOLLOWING ENV, ΔSLOPE, SLOPECOMP

Not submitted to run: R^2, DUR, PAUSEN, PAUSEDUR, ARTRATE, SPKRATE

### 10.4.4. Interpreting the results

The results here, though putative (see the next section), appear quite promising. For the two variables examined, (ð) and (ing), slope-based metrics from the Henderson graph arise as important independent variables contributing to our understanding in the factors behind the sociolinguistic variation.
(ing) patterns as we might expect (cf. Fischer 1958; Hazen 2008). The older males realize –*in’ most of the time, while the younger females have more variable productions. The fact that steeper slopes are found to predict velar, –*ing, realizations supports the notion that (ing) is affected by attention to speech, or formality (Fischer 1958; Labov 1966, 2001), and – more importantly for our present purposes – that the slopes of the Henderson graph segments do in fact appear to relate to the attentiveness of the speaker to his or her own speech.

(ð), the variable of greater interest for the Petty Harbour data, also bears a relationship to the Henderson graph’s characterization of the sequential temporal patterns of speech. However, this relationship is less straightforward than for (ing). (ð) is found here to relate for the young females to changes in the Henderson slope between adjacent segments. Positive changes – that is, changes towards more hesitant speech – are found to predict more likely realization of stopped (ð), [d], while changes towards more fluent speech – i.e. shallower slopes or negative changes – predict more fricated realizations, [ð]. This was illustrated in Figure 10.4.2.

Importantly, the relationship between the character of the slope and the realization of (ð) are only significant for the young females. The old males do not show a relationship with any temporal sequential variable. This is good evidence in support of the interpretation that the variable (ð) is grammaticalized differently for the young females than for the old males.17 In short, it strengthens our interpretation that this

17 Of course, we still expect that some linguistic factors would arise as predictors for the old males’ use of (ð). The limitations of the data here – only examining five similar lexical types – restricts us from examining wider sets of linguistic features (cf. Van Herk et al. forthcoming).
community is undergoing language change. It is also strong evidence in support of the interpretation that (ð) is a salient variable in Petty Harbour (Van Herk et al. forthcoming) and that the young females actively, or “consciously” in Chafe’s (1994) terms, make use of the variable. We also find that lexical type is an important factor in (ð) realization for the young females. This indicates further that certain lexical items are more available than others for performative, identity-related work.

To briefly return to the question I posed earlier of how we can choose between two conflicting interpretations for the use of (ð) in Petty Harbour (i.e. do the young speakers suppress or enact stopped (ð)?), the evidence from the sequential temporal measures seems to indicate that they enact the marked local form ([d] for (ð)). If we believe Henderson et al. (1966) that steeper slopes indicate more hesitant speech, and we believe, or are willing to humor, the equation of hesitant speech with more “conscious” speech (Chafe 1994) or more attentive speech (Labov 1966, 1972), the more “performed” or “enacted” or “consciously-chosen” variable form should correlate with the steeper slopes, and that is the case for [d], the stopped variant of (ð) in this analysis. The Henderson data seem incongruous with an interpretation of the young females actively suppressing the marked local form for the data examined here. The fact that this conflicts with Van Herk et al.’s (forthcoming) argument and some compelling comparative evidence (Van Herk personal communication) will have to remain an issue to be resolved at a later date.

---

18 As I believe we should based on at least the analysis in §10.5.3 of (ing).
Regardless of which understanding we adopt, we have shown here that there is indeed a difference between the locally salient and non-salient variables, thus verifying the claim for the systematic relationship between paralinguistic cues and marked dialect forms. More generally, the Henderson graph method appears to provide promising new quantitative metrics to paralinguistic cues to speech style (cf. Labov 1972).

10.4.5. Mitigating the results

The case study presented in this chapter is intended to be illustrative and putative more than conclusive. There are a number of factors (beyond the conflict with the interpretation made by Van Herk et al. forthcoming) that mitigate the present results and are worth discussing further.

While the Henderson slope lines are generated using a least-squares estimate of best-fit, the beginnings and ends of the slope lines – i.e. the particular spans of talk that are segmented within a single slope line – are produced by analyst inspection. This is necessarily interpretive and subjective work. Nonetheless, I based my segmentation on a principled determination from a combination of the Henderson graph’s characteristics and cues from the discourse content itself. Henderson et al. (1966) made use of this “inspection” approach and appear to have been pleased with its success. They state that
some “passages were not as clear, nor the lines as confidently fitted. However, the same alternating pattern was quite discernible in all the spontaneous speech passages” (208).19

Readers may also have noted that the data examined here seem quite appropriate to analyze using mixed-effect models, as was done for the data in Chapters 5 and 6. Despite the fact that the data here, indeed, appear to be appropriately considered as multi-level data, with individual speakers as random effects, I have presented the analyses as fixed-effects logistic regressions. In simplistic terms, the output of mixed-effects models are primarily more conservative than fixed-effect models. Preliminary analyses using mixed models failed to obtain significance for the temporal sequencing variables.20 I nonetheless chose in this work to present the fixed-effects analyses since these analyses speak to the questions of interest here.21 More data and further statistical analyses are required in order to determine whether the differences in the statistical models are consequences of the case study-like format of this analysis (with only seven speakers and 1.6 hours of analyzed talk) or are indicating a larger problem for the Henderson-based

19 I have experimented with less subjective methods that might allow the determination of the segmentation for the slope lines based on automatic approaches. For example, given a stretch of (uninterrupted) talk by a speaker, an algorithm could take a step-up or step-down approach, comparing the correlation coefficients ($R^2$ values) of every possible combination of slopes lines and choosing the best set of segments based on the best set of correlation coefficients. The criteria for determining the “best set” of correlation coefficients, however, is complex and requires more work than I have been able to devote to here. Moreover, having manually selected the Henderson segments for the entire 1.6 hours of speech examined here, I am skeptical of the ability of such an automated procedure. In delimiting segments, I have used not only the visual character of the Henderson graphs, but also the discourse content to identify meaningful changes in the character of the talk. I am skeptical that an automated procedure could do as appropriate a job of interpreting the character of the talk, regardless of how well it may be able to fit lines to the overall shape of the plot.

20 Note that these variables are significant in two of the models above with $p$ values above 0.01. A slightly more conservative model could easily push these values above 0.05, no longer in the realm of significance.

21 I also rationalized this decision by considering the ubiquity of fixed-effect logistic (Varbrul-like) regression in sociolinguistics. While I am convinced of the benefits of mixed-effect analysis for sociolinguistic work (and hence used mixed-effect methods for Chapters 5 and 6), I believe that it is congruous with generally accepted sociolinguistic practice to use fixed-effect models when necessary if they help highlight relevant patterns within the data.
metrics. At present, I feel confident that more thorough mixed-effect models will obtain similar results once more speakers and speech data are added to the analysis.

A final mitigating factor comes from the obvious sense that there are surely combinations of factors that influence the slope beyond “consciousness” or attention to speech or cognitive activity. A finer-grained analysis using this metric will need to account for other contextual factors (such as environmental distractions and non-verbal components of communication) to fully understand the relationship between a set of Henderson slope measures and the actual parameters of the discourse.

10.5. In closing: sequential temporal patterns and variable realization

While the previous chapters of Part 2 have examined speech rate and pause for their own sake – with the principal goals of understanding how speech rate and pause pattern when considered as sociolinguistic variables – this chapter has shown that speech rate and pause are also useful when it comes to understanding the realization of more “typical” sociolinguistic variables. Through this consideration we have developed a new quantitative metric that appears quite promising for assessing the relationship between paralinguistic cues and variable realization. By bringing the productive tradition of psycholinguistics’ consideration of pause (and speech rate) as indicators of cognitive activity to bear on the notion of paralinguistic cues and performance, we have also, I hope, broadened the scope of sociolinguistic studies of style, at least in the attention to speech model, to include an interest in speakers’ processes of production.
Part 3. Towards the future

11. Conclusion

11.1. In final closing

This dissertation has attempted to balance dual goals, to argue for a particular approach to sociolinguistic data and its management (Part 1) and to evaluate variation in speech rate and pause, and its social and linguistic correlates (Part 2). In having two goals, it has not perhaps cohered as nicely as some dissertations do. Nonetheless, I hope to have shown in these pages that new ways of conceptualizing collections of sociolinguistic data can enrich sociolinguistic research and sociolinguistic research groups. Bringing together the many recordings made over the course of diverse projects and creating a centralized archive with an explicit and extensible annotation scheme builds richer data sets, and this enriched data can be used to extend questions of sociolinguistic variation beyond the morphosyntactic and phonological on the one hand and the sociophonetic on the other.

As shown in Part 2, these new approaches to the analysis of language variation shed light on a wide range of questions and problems for sociolinguistics. From even just the exploratory analyses presented here, we have learned a great deal about the possibilities of language variation and its potential meanings. In the next section I
enumerate the various substantive findings of Part 2. However, more than any particular finding, I believe this work has shown that language variation pervades all aspects of speech – even levels as minute as the quantitative differences found here in pause durations and speech rate measurements – and that this variation is multidimensional – existent within and between individual speakers, speech communities and demographic characterizations, and interactions between speakers and speakers’ collaborative and conflicting presentations of identity.

11.2. Summary of findings (from Part 2)

Part 2 of this dissertation presented a number of findings on variation in speech rate and pause. I here briefly summarize the most important of these.

Chapter 5, pause: From the 22,734 pause measurements from 104 speakers examined here, the overall finding on pause is that pause realization, in addition to being an outcome of cognitive processes, is socially impacted. Specifically, region, gender, and ethnicity are all found to be significant factors in the durations of silent pauses. The African American teenagers from Washington, DC (the only speakers examined here from Washington, DC) favor long pauses. Speakers from Western NC slightly favor longer pauses, while speakers from other parts of North Carolina and from Texas are found to slightly favor shorter pauses. The Ohioans in the dataset, by far, favor short pauses. While gender and ethnicity are found to be significant, they do not yield as large
a range of variation as region. Men appear to have slightly longer pauses than women, and Lumbees and Latinos/as are found to slightly favor longer pauses than European Americans and African Americans.

All in all, the variation in pause realization exhibited by the 104 speakers examined here is argued to be meaningful in the sense that the differences between speakers are likely perceivable by listeners and, thus, this variation plays a role in speakers self-presentation and hearers perceptions of other talkers.

**Chapter 6, speech rate:** The analysis of the 23,871 tokens of speech rate data for 104 speakers show that social patterns in speech rate, on the one hand, are strong enough to be readily noticeable from simple descriptive statistics of the data’s distribution (e.g., Figure 6.5.2), but that, on the other hand, complex interactions of the gender and ethnicity of the speakers complicate a simplistic explanation of the role of any one demographic factor. Meanwhile, the picture of speech rate variation is further complicated by the fact that speech rate shows a very strong linguistic effect of utterance length (in terms of syllables per utterance). Including utterance length in a statistical model for the speech rate data yields a picture of speech rate variation where the combined, interacting gender and ethnicity of the speaker is the only significant social category. Without considering utterance length, region is found to be highly significant, with roughly the inverse order found in the previous chapter for pause duration. Ethnicity is also found to be significant, again with the inverse order found for pause.
Depending on the specific model, gender and age are found to play significant roles in predicting speech rate, though the effect of age is at best minor.

Chapter 7, pause and speech rate: Here I considered and investigated more deeply the observation that pause and speech rate were found to have similar, though inverse, outcomes in their individual regression models. The analysis of the combined data found that pause and speech rate do, in fact, arise as significant (inverse) predictors for one another. This predictive relationship is not found to be as strong, however, as the social effects, especially of region. Moreover, we find that for the Lumbees, but only the Lumbees, the most tightly knit “community” examined here, there is a very strong inverse linear relationship between pause and speech rate. I presented some putative explanations for why that clear pattern is found and only found for the Lumbees.

Chapter 8, style and register: Here I briefly assessed the role that macro-level conceptions of style (on par with register variation) play on the realization of speech rate and pause by a case study examining three public figures across two different speech settings each. I found that speech rate and pause can be manifested differently by different speakers under different settings, but that the same is found to be true for “normal” sociolinguistic variables. Overall, though, as we might expect from the psycholinguistic literature, speech rate and pause do exhibit sensitivity to register variation and we find longer pauses or slower speech rates correlating with “read-like” registers in comparison to “interview speech” registers.
Chapter 9, interaction effects: Here I approached the extent of interactional differences in the data, both from macro-level, quantitative perspectives relating to the gender, ethnicity, and number of participants in each interaction and from a more micro-level conception of style through two case studies each examining intra-speaker variation by the same individuals across multiple interviews. In terms of the effects of participants, I found perhaps the strongest factor in all of the data examined: Female interviewers elicit the slowest speech rates (regardless of the gender of the interviewees) and, in fact, models that include the gender of the interviewer (e.g., Table 9.3.2) find only interviewer gender to be a significant predictor for speech rate. I also found other effects of the ethnicity of the interviewer and the number of participants on speech rate and slight indication that there is an effect of interviewer gender on pause, but this interviewer gender effect on speech rate is by far the most striking.

The case studies in Chapter 9, one on an interviewee across five interviews and one on an interviewer across 12 interviews with different interviewees, show that speakers can be differently attuned to the speech of their interlocutors and may or may not accommodate aspects of their speech rate and pause with those of their interlocutors. I also suggested that interviewers may be more attentive and do more accommodating than interviewees, an idea that seems intuitively reasonable but is also supported by the two case studies.
Chapter 10, pause and speech rate as paralinguistic cues to other variables:

In the final chapter of Part 2, I took a more conventional view of speech rate and pause and developed a metric based on the *Henderson graph* (§3.4.1; Henderson et al. 1966) to quantitatively assess the status of speech rate and pause (and more generally a class of measures termed *sequential temporal variables*) as paralinguistic cues to attention to speech. Through this, I found that a combinative measure of sequential temporal patterning – the *Henderson slope* – correlates changes to greater speech hesitancy (or increased cognitive processing) with the use of a locally marked variable feature, the stopping of (voiced) interdental fricatives. I also find the validity of the slope metric is confirmed through its correlating with variation in velar nasal fronting, which has been long established to be effected by formality. Steeper slopes – again, indicating more hesitant speech – correlate with higher likelihood of full velar nasal productions.

11.3. For the future

I end this dissertation with a section titled “For the future” since it will be many years before we know the true impact of SLAAP on the field of sociolinguistics. The ultimate contribution of SLAAP – more than any improved conceptualization or practice of data management, annotation, and so forth – may prove simply to be its ability to preserve and make accessible sociolinguistic recordings for the future and, through the Internet, across geographical distance. While this has rarely been my central goal, if it proves to be the case, it will more than justify the many hours (nay, years) that the
members of the NCLLP and I have put into its development and the development of its
archive. Nonetheless, work on SLAAP is ongoing, both in terms of the growth of the
archive and in terms of the utility of the software, and I look forward to the future
directions that SLAAP takes. There is plenty of work still to do.

For myself, this dissertation has inspired a larger project in linguistics that I might
call “social psycholinguistics”. I am convinced of the importance of empirical analyses
of real-world talk and of the importance of improved understandings of the principles and
bounds of linguistic variation for linguistics’ broader pursuit of the nature of language.
However, I also believe that sociolinguistics – with its rich empirical toolkit and data
collection methods – has more to offer directly to our understanding of language from
cognitive perspectives and formal perspectives than it often contributes. I hope this work
has illustrated one small way that we can move towards a more unified empirical
understanding of language, where variation can be studied as both a symptom of and
window into language and cognitive processing.
### Appendix: Data summary for primary 104 speakers used in Part 2, ordered by year of birth

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Ethnicity</th>
<th>Gender</th>
<th>YOB</th>
<th>Age</th>
<th>Geography</th>
<th>Region</th>
<th>SR N</th>
<th>SR Med</th>
<th>SR Mean</th>
<th>SR SD</th>
<th>P N</th>
<th>P Med</th>
<th>P Mean</th>
<th>P SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEL</td>
<td>Lumbee</td>
<td>Male</td>
<td>1881</td>
<td>90</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>72</td>
<td>4.54</td>
<td>4.44</td>
<td>1.07</td>
<td>76</td>
<td>496</td>
<td>637</td>
<td>572</td>
</tr>
<tr>
<td>FA</td>
<td>Black</td>
<td>Female</td>
<td>1908</td>
<td>87</td>
<td>Warren</td>
<td>Central NC</td>
<td>92</td>
<td>4.63</td>
<td>4.55</td>
<td>1.47</td>
<td>65</td>
<td>314</td>
<td>472</td>
<td>426</td>
</tr>
<tr>
<td>MMC</td>
<td>White</td>
<td>Male</td>
<td>1910</td>
<td>86</td>
<td>Hyde</td>
<td>Eastern NC</td>
<td>224</td>
<td>5.27</td>
<td>5.22</td>
<td>1.48</td>
<td>277</td>
<td>524</td>
<td>636</td>
<td>490</td>
</tr>
<tr>
<td>RS</td>
<td>Black</td>
<td>Male</td>
<td>1910</td>
<td>86</td>
<td>Hyde</td>
<td>Eastern NC</td>
<td>132</td>
<td>4.33</td>
<td>4.29</td>
<td>1.09</td>
<td>93</td>
<td>333</td>
<td>456</td>
<td>405</td>
</tr>
<tr>
<td>JAA</td>
<td>White</td>
<td>Male</td>
<td>1915</td>
<td>80</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>62</td>
<td>4.64</td>
<td>4.69</td>
<td>1.23</td>
<td>65</td>
<td>286</td>
<td>388</td>
<td>304</td>
</tr>
<tr>
<td>Karen</td>
<td>Black</td>
<td>Female</td>
<td>1916</td>
<td>86</td>
<td>Texana</td>
<td>Western NC</td>
<td>96</td>
<td>3.66</td>
<td>3.43</td>
<td>1.14</td>
<td>83</td>
<td>311</td>
<td>435</td>
<td>444</td>
</tr>
<tr>
<td>JB</td>
<td>Lumbee</td>
<td>Male</td>
<td>1918</td>
<td>75</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>111</td>
<td>4.75</td>
<td>4.61</td>
<td>1.18</td>
<td>72</td>
<td>422</td>
<td>625</td>
<td>594</td>
</tr>
<tr>
<td>MP</td>
<td>Latino/a</td>
<td>Female</td>
<td>1920</td>
<td>87</td>
<td>Pearsall</td>
<td>Texas</td>
<td>172</td>
<td>4.38</td>
<td>4.42</td>
<td>1.31</td>
<td>205</td>
<td>570</td>
<td>733</td>
<td>611</td>
</tr>
<tr>
<td>AD</td>
<td>Lumbee</td>
<td>Male</td>
<td>1921</td>
<td>50</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>26</td>
<td>4.34</td>
<td>4.15</td>
<td>1.42</td>
<td>22</td>
<td>482</td>
<td>888</td>
<td>1146</td>
</tr>
<tr>
<td>LB</td>
<td>Lumbee</td>
<td>Female</td>
<td>1923</td>
<td>70</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>117</td>
<td>4.75</td>
<td>4.74</td>
<td>1.22</td>
<td>101</td>
<td>411</td>
<td>527</td>
<td>423</td>
</tr>
<tr>
<td>EYsr</td>
<td>White</td>
<td>Male</td>
<td>1923</td>
<td>81</td>
<td>Pearsall</td>
<td>Texas</td>
<td>129</td>
<td>5.59</td>
<td>5.53</td>
<td>1.18</td>
<td>280</td>
<td>474</td>
<td>503</td>
<td>302</td>
</tr>
<tr>
<td>CMcK</td>
<td>White</td>
<td>Male</td>
<td>1923</td>
<td>84</td>
<td>Pearsall</td>
<td>Texas</td>
<td>73</td>
<td>4.82</td>
<td>4.81</td>
<td>0.80</td>
<td>58</td>
<td>540</td>
<td>672</td>
<td>475</td>
</tr>
<tr>
<td>EH</td>
<td>Black</td>
<td>Female</td>
<td>1924</td>
<td>82</td>
<td>Wilson</td>
<td>Western NC</td>
<td>922</td>
<td>4.08</td>
<td>4.02</td>
<td>1.15</td>
<td>811</td>
<td>395</td>
<td>434</td>
<td>346</td>
</tr>
<tr>
<td>LMcK</td>
<td>White</td>
<td>Female</td>
<td>1925</td>
<td>81</td>
<td>Pearsall</td>
<td>Texas</td>
<td>50</td>
<td>4.80</td>
<td>4.67</td>
<td>0.98</td>
<td>40</td>
<td>343</td>
<td>418</td>
<td>369</td>
</tr>
<tr>
<td>CV</td>
<td>Latino/a</td>
<td>Female</td>
<td>1926</td>
<td>78</td>
<td>Pearsall</td>
<td>Texas</td>
<td>70</td>
<td>5.31</td>
<td>5.07</td>
<td>1.23</td>
<td>92</td>
<td>328</td>
<td>495</td>
<td>507</td>
</tr>
<tr>
<td>ALPD</td>
<td>White</td>
<td>Female</td>
<td>1926</td>
<td>84</td>
<td>Warren</td>
<td>Central NC</td>
<td>148</td>
<td>4.83</td>
<td>4.88</td>
<td>1.31</td>
<td>139</td>
<td>547</td>
<td>683</td>
<td>483</td>
</tr>
<tr>
<td>EKM</td>
<td>Black</td>
<td>Male</td>
<td>1927</td>
<td>67</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>117</td>
<td>5.22</td>
<td>5.03</td>
<td>1.55</td>
<td>69</td>
<td>479</td>
<td>526</td>
<td>352</td>
</tr>
<tr>
<td>ALL</td>
<td>White</td>
<td>Female</td>
<td>1927</td>
<td>68</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>135</td>
<td>5.28</td>
<td>5.07</td>
<td>1.24</td>
<td>134</td>
<td>372</td>
<td>436</td>
<td>335</td>
</tr>
<tr>
<td>DMcK</td>
<td>White</td>
<td>Male</td>
<td>1927</td>
<td>79</td>
<td>Pearsall</td>
<td>Texas</td>
<td>48</td>
<td>4.50</td>
<td>4.60</td>
<td>1.13</td>
<td>116</td>
<td>445</td>
<td>514</td>
<td>328</td>
</tr>
<tr>
<td>LJ</td>
<td>Black</td>
<td>Female</td>
<td>1928</td>
<td>66</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>97</td>
<td>4.32</td>
<td>4.27</td>
<td>1.36</td>
<td>52</td>
<td>326</td>
<td>386</td>
<td>334</td>
</tr>
<tr>
<td>HS</td>
<td>White</td>
<td>Male</td>
<td>1929</td>
<td>77</td>
<td>Pearsall</td>
<td>Texas</td>
<td>82</td>
<td>5.17</td>
<td>5.24</td>
<td>1.09</td>
<td>60</td>
<td>487</td>
<td>672</td>
<td>504</td>
</tr>
<tr>
<td>GA</td>
<td>Black</td>
<td>Male</td>
<td>1930</td>
<td>64</td>
<td>Warren</td>
<td>Central NC</td>
<td>76</td>
<td>5.24</td>
<td>5.19</td>
<td>1.24</td>
<td>94</td>
<td>314</td>
<td>412</td>
<td>353</td>
</tr>
<tr>
<td>DJC</td>
<td>White</td>
<td>Female</td>
<td>1930</td>
<td>76</td>
<td>Pearsall</td>
<td>Texas</td>
<td>133</td>
<td>4.89</td>
<td>4.89</td>
<td>1.30</td>
<td>151</td>
<td>311</td>
<td>384</td>
<td>318</td>
</tr>
<tr>
<td>Gaill</td>
<td>Black</td>
<td>Female</td>
<td>1932</td>
<td>70</td>
<td>Texana</td>
<td>Western NC</td>
<td>836</td>
<td>4.28</td>
<td>4.25</td>
<td>1.31</td>
<td>576</td>
<td>360</td>
<td>502</td>
<td>466</td>
</tr>
<tr>
<td>GL</td>
<td>Lumbee</td>
<td>Female</td>
<td>1937</td>
<td>56</td>
<td>Robeson</td>
<td>Southern NC</td>
<td>66</td>
<td>4.34</td>
<td>4.28</td>
<td>1.41</td>
<td>56</td>
<td>500</td>
<td>787</td>
<td>805</td>
</tr>
<tr>
<td>RC</td>
<td>Latino/a</td>
<td>Male</td>
<td>1937</td>
<td>66</td>
<td>Pearsall</td>
<td>Texas</td>
<td>70</td>
<td>4.58</td>
<td>4.69</td>
<td>1.25</td>
<td>82</td>
<td>418</td>
<td>593</td>
<td>475</td>
</tr>
<tr>
<td>CLS</td>
<td>White</td>
<td>Male</td>
<td>1937</td>
<td>69</td>
<td>Pearsall</td>
<td>Texas</td>
<td>118</td>
<td>5.08</td>
<td>5.14</td>
<td>1.35</td>
<td>145</td>
<td>452</td>
<td>637</td>
<td>601</td>
</tr>
<tr>
<td>IG</td>
<td>Latino/a</td>
<td>Female</td>
<td>1938</td>
<td>68</td>
<td>Pearsall</td>
<td>Texas</td>
<td>140</td>
<td>4.60</td>
<td>4.49</td>
<td>1.34</td>
<td>199</td>
<td>407</td>
<td>501</td>
<td>374</td>
</tr>
<tr>
<td>ORH</td>
<td>White</td>
<td>Male</td>
<td>1938</td>
<td>69</td>
<td>Beech</td>
<td>Western NC</td>
<td>541</td>
<td>5.49</td>
<td>5.50</td>
<td>1.48</td>
<td>488</td>
<td>538</td>
<td>780</td>
<td>713</td>
</tr>
<tr>
<td>Commis-</td>
<td>Black</td>
<td>Female</td>
<td>1941</td>
<td>61</td>
<td>Roanoke</td>
<td>Eastern NC</td>
<td>89</td>
<td>3.92</td>
<td>3.93</td>
<td>1.01</td>
<td>97</td>
<td>371</td>
<td>525</td>
<td>550</td>
</tr>
</tbody>
</table>

240
<table>
<thead>
<tr>
<th>County, NC</th>
<th>Race</th>
<th>Sex</th>
<th>Year</th>
<th>Age</th>
<th>Location</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robeson</td>
<td>Black</td>
<td>Male</td>
<td>1972</td>
<td>22</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>Warren</td>
<td>Black</td>
<td>Male</td>
<td>1978</td>
<td>20</td>
<td>Princeville</td>
<td>NC</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1978</td>
<td>50</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>NC</td>
<td>Female</td>
<td></td>
<td>1977</td>
<td>37</td>
<td>Princeville</td>
<td>NC</td>
</tr>
<tr>
<td>Princeville</td>
<td>Female</td>
<td></td>
<td>1978</td>
<td>30</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Raleigh</td>
<td>Black</td>
<td>Male</td>
<td>1976</td>
<td>31</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1977</td>
<td>35</td>
<td>Hudson, OH</td>
<td>NC</td>
</tr>
<tr>
<td>Robeson</td>
<td>Black</td>
<td>Male</td>
<td>1972</td>
<td>20</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1978</td>
<td>50</td>
<td>Princeville</td>
<td>NC</td>
</tr>
<tr>
<td>NC</td>
<td>Female</td>
<td></td>
<td>1977</td>
<td>37</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Princeville</td>
<td>Female</td>
<td></td>
<td>1978</td>
<td>30</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Raleigh</td>
<td>Black</td>
<td>Male</td>
<td>1976</td>
<td>31</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1977</td>
<td>35</td>
<td>Hudson, OH</td>
<td>NC</td>
</tr>
<tr>
<td>Robeson</td>
<td>Black</td>
<td>Male</td>
<td>1972</td>
<td>20</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1978</td>
<td>50</td>
<td>Princeville</td>
<td>NC</td>
</tr>
<tr>
<td>NC</td>
<td>Female</td>
<td></td>
<td>1977</td>
<td>37</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Princeville</td>
<td>Female</td>
<td></td>
<td>1978</td>
<td>30</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Raleigh</td>
<td>Black</td>
<td>Male</td>
<td>1976</td>
<td>31</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1977</td>
<td>35</td>
<td>Hudson, OH</td>
<td>NC</td>
</tr>
<tr>
<td>Robeson</td>
<td>Black</td>
<td>Male</td>
<td>1972</td>
<td>20</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1978</td>
<td>50</td>
<td>Princeville</td>
<td>NC</td>
</tr>
<tr>
<td>NC</td>
<td>Female</td>
<td></td>
<td>1977</td>
<td>37</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Princeville</td>
<td>Female</td>
<td></td>
<td>1978</td>
<td>30</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Raleigh</td>
<td>Black</td>
<td>Male</td>
<td>1976</td>
<td>31</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1977</td>
<td>35</td>
<td>Hudson, OH</td>
<td>NC</td>
</tr>
<tr>
<td>Robeson</td>
<td>Black</td>
<td>Male</td>
<td>1972</td>
<td>20</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1978</td>
<td>50</td>
<td>Princeville</td>
<td>NC</td>
</tr>
<tr>
<td>NC</td>
<td>Female</td>
<td></td>
<td>1977</td>
<td>37</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Princeville</td>
<td>Female</td>
<td></td>
<td>1978</td>
<td>30</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Raleigh</td>
<td>Black</td>
<td>Male</td>
<td>1976</td>
<td>31</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1977</td>
<td>35</td>
<td>Hudson, OH</td>
<td>NC</td>
</tr>
<tr>
<td>Robeson</td>
<td>Black</td>
<td>Male</td>
<td>1972</td>
<td>20</td>
<td>Pearsall</td>
<td>TX</td>
</tr>
<tr>
<td>NC</td>
<td>Male</td>
<td></td>
<td>1978</td>
<td>50</td>
<td>Princeville</td>
<td>NC</td>
</tr>
<tr>
<td>NC</td>
<td>Female</td>
<td></td>
<td>1977</td>
<td>37</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Princeville</td>
<td>Female</td>
<td></td>
<td>1978</td>
<td>30</td>
<td>Wilmington</td>
<td>NC</td>
</tr>
<tr>
<td>Name</td>
<td>Gender</td>
<td>Year of Birth</td>
<td>County or City</td>
<td>Age</td>
<td>Race</td>
<td>Ethnicity</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>---------------</td>
<td>-----------------</td>
<td>-----</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Lori</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Jalinda</td>
<td>Female</td>
<td>1994</td>
<td>Durham, NC</td>
<td>125</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Elizabeth</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Celso</td>
<td>Male</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Marcela E.</td>
<td>Female</td>
<td>1994</td>
<td>Durham, NC</td>
<td>125</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Catrina</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Cleo</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Elizabeth</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Jalinda</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Lori</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Marcela E.</td>
<td>Female</td>
<td>1994</td>
<td>Durham, NC</td>
<td>125</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Catrina</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Cleo</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Elizabeth</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Jalinda</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
<tr>
<td>Lori</td>
<td>Female</td>
<td>1995</td>
<td>Durham, NC</td>
<td>126</td>
<td>Black</td>
<td>Latino/a</td>
</tr>
</tbody>
</table>

242
<table>
<thead>
<tr>
<th>Name</th>
<th>Gender/Origin</th>
<th>Sex</th>
<th>Year</th>
<th>Month</th>
<th>City</th>
<th>Region</th>
<th>Age</th>
<th>Height</th>
<th>Weight</th>
<th>BP1</th>
<th>BP2</th>
<th>BP3</th>
<th>BMI</th>
<th>Weight</th>
<th>Ht</th>
<th>Time</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rolando</td>
<td>Latino/a</td>
<td>Male</td>
<td>1995</td>
<td>12</td>
<td>Durham, NC</td>
<td>Central NC</td>
<td>75</td>
<td>4.34</td>
<td>4.13</td>
<td>1.33</td>
<td>77</td>
<td>501</td>
<td>808</td>
<td>859</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silvia</td>
<td>Latino/a</td>
<td>Female</td>
<td>1995</td>
<td>12</td>
<td>Durham, NC</td>
<td>Central NC</td>
<td>237</td>
<td>3.14</td>
<td>3.03</td>
<td>1.11</td>
<td>236</td>
<td>360</td>
<td>508</td>
<td>449</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Che</td>
<td>Latino/a</td>
<td>Male</td>
<td>1996</td>
<td>11</td>
<td>Durham, NC</td>
<td>Central NC</td>
<td>181</td>
<td>3.92</td>
<td>3.89</td>
<td>1.36</td>
<td>228</td>
<td>529</td>
<td>778</td>
<td>737</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>Latino/a</td>
<td>Male</td>
<td>1997</td>
<td>10</td>
<td>Pearsall, TX</td>
<td>Texas</td>
<td>84</td>
<td>5.07</td>
<td>5.16</td>
<td>1.30</td>
<td>83</td>
<td>363</td>
<td>387</td>
<td>272</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Cooper, Todd. 2007. Documenting the American South. Paper presented in special panel, Re-examining Language Data in the American South, at the South Eastern Conference on Linguistics (SECOL) 74: Natchitoches, LA.


Mukherjee, Joybrato. 2000. Speech is silver, but silence is golden: Some remarks on the function(s) of pauses. *Anglia* 118: 571-84.


Biography

Tyler Kendall was born in Hanover, New Hampshire, on September 16th, 1976. In 1977 he moved with his parents and older brother to Martha’s Vineyard, Massachusetts, where he attended public school at the West Tisbury Elementary School and then the Martha’s Vineyard Regional High School. He left Martha’s Vineyard in 1994 to attend college at Cornell University, in Ithaca, New York. Tyler graduated Cornell in 1998 with B.A. degrees in Archaeology and Classics. In March of 2000, work with an Internet start up brought him to Raleigh, North Carolina. He soon found himself beginning a long and ongoing relationship with North Carolina State University. From 2001 to 2003, he worked in the Information Technology Division and Disability Services for Students offices at NCSU, first as the University’s first Web Accessibility Specialist and then as its second Coordinator of Assistive and Information Technology. It was in those roles that he first met Erik Thomas and Walt Wolfram and discovered sociolinguistics. Soon after, he left his employment at NCSU to pursue a Ph.D. in sociolinguistics – an adventure that ends with the following page.
