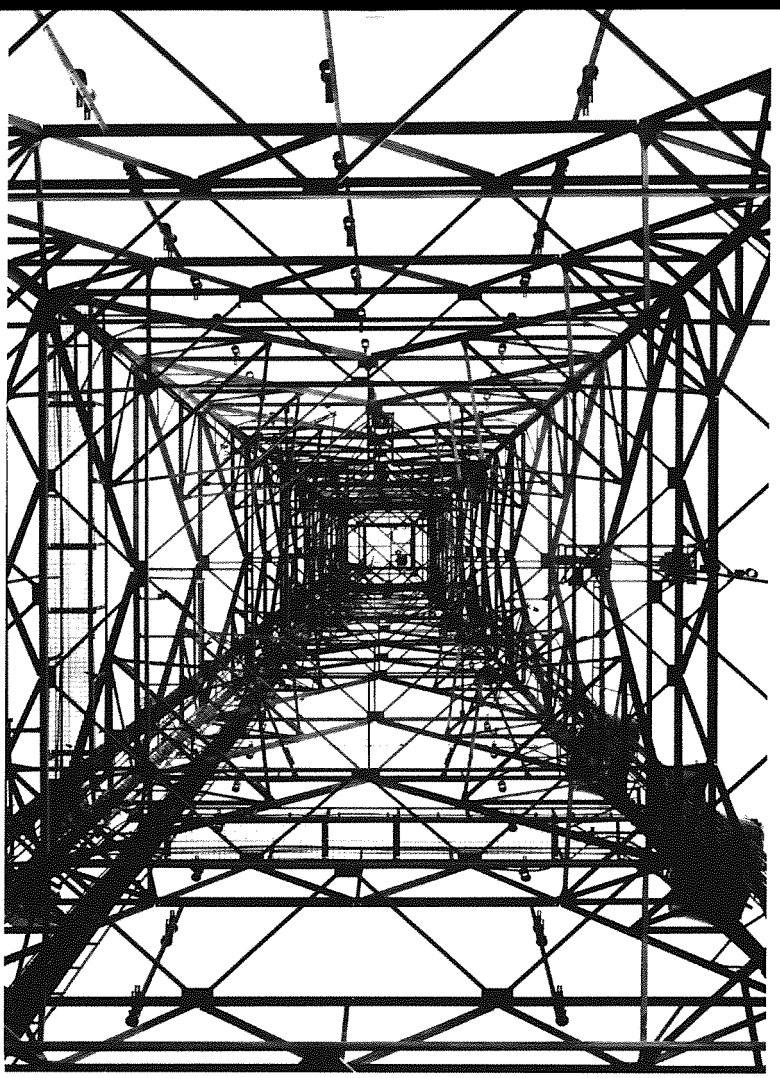


SPEECH RATE, PAUSE, AND SOCIOLINGUISTIC VARIATION

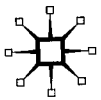
STUDIES IN CORPUS SOCIOPHONETICS

TYLER KENDALL

Speech Rate, Pause, and Sociolinguistic Variation

Studies in Corpus Sociophonetics

Tyler Kendall
University of Oregon, USA



First published 2013 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978–0–230–24977–6

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1
22 21 20 19 18 17 16 15 14 13

Printed and bound in Great Britain by
CPI Antony Rowe, Chippenham and Eastbourne

Contents

List of Figures viii

List of Tables xi

Acknowledgments xiii

Part I Speech Rate, Pause, and Corpus Sociophonetics

1 Looking Forward 3
- 1.1 Introduction 3
- 1.2 Disciplinarity and intersections 5
- 1.3 Why exactly speech rate and pause? 8
- 1.4 Overview of the monograph 10

2 What We Know about Speech Rate and Pause 12
- 2.1 Introduction 12
- 2.2 Attitudes towards and the perception of speech rate and pause 14
- 2.3 Pauses in detail 20
- 2.4 Speech rates in detail 26
- 2.5 Motivating further study 35

3 New Tools and Speech Databases 37
- 3.1 Introduction 37
- 3.2 The Sociolinguistic Archive and Analysis Project (SLAAP) 38
- 3.3 SLAAP's transcript model 40
- 3.4 The Online Speech/Corpora Archive and Analysis Resource 44
- 3.5 Tools for the analysis of temporal speech features 45

Part II Studies in Speech Rate and Pause Variation

4 Methods and a First Look at Speech Rate and Pause 51
- 4.1 Introduction 51
- 4.2 Modeling sociophonetic data 52
- 4.3 The reading passage data 56
- 4.4 Measuring and defining rate of speech and pause 58
  - 4.4.1 Rate of speech 58
  - 4.4.2 Pause durations 63

4.5 Reading passage data and analysis 64
4.5.1 Rate of speech in the reading passage data and its statistical analysis 66
4.5.2 Pauses in the reading passage data 79
4.6 From investigating read data to conversational speech data 80

5 Speech Rate and Pause in Conversational Interviews 83
5.1 Introduction 83
5.2 The data 84
5.3 Modeling speech rate and pause durations at the measurement level 89
5.3.1 Speech rate at the utterance level 90
5.3.2 Pause duration at the pause level 97
5.4 Modeling speech rate and pause durations at the speaker level 101
5.4.1 Speech rate at the speaker level 102
5.4.2 Pause duration at the speaker level 109
5.5 Which approach is better? 115
5.6 The sociolinguistic patterns of speech rate and pause duration 117

6 Closer Looks at Speech Rate and Pause Variation: Methods and Findings 121
6.1 Introduction 121
6.2 How many speech rate measurements yield stable patterns? 122
6.2.1 The stability of central tendencies 123
6.2.2 Measurement size and the stability of the statistical models 125
6.2.3 Making sense of conflicting results 129
6.3 How long is a pause? (An experiment in modeling) 130
6.4 Articulation rates in Intonational Phrases and the effect of phrase-final lengthening 138
6.5 Pause duration variability as a function of pause type 148
6.6 Summing up 156

7 Closer Looks at Speech Rate and Pause Variation: Interlocutors and Accommodation 158
7.1 Introduction 158
7.2 Interlocutor effects on speech rate and pause 159

7.3 Accommodation in pauses and speech rates 167
7.3.1 A case study: who is interviewing EH? 167
7.3.2 A case study: C is interviewing whom? 170
7.4 Summing up 176

Part III Speech Rate, Pause, and Sociolinguistic Variation

8 The Influence of Speech Rate and Pause on Sociolinguistic Variables 181
8.1 Introduction 181
8.2 The sociolinguistics of style 184
8.3 The psycholinguistics of style 186
8.4 Channel cues to attention to speech 188
8.5 The Henderson graph: a method for quantifying attention to speech 190
8.5.1 A new methodology for Henderson graphing 193
8.5.2 Henderson graph-based metrics 196
8.6 Case study: the interviews with adolescent African American girls in Washington, DC 197
8.6.1 Henderson graph slopes and sequential temporal variation 197
8.6.2 Hesitancy in narrative versus nonnarrative talk 199
8.6.3 Attention to speech and variable (ing) 200
8.6.4 Channel cues in the DC interviews 206
8.7 Conclusion 206

9 Looking Back and Looking Further Forward 210
9.1 Taking stock 210

Appendix I: Guide to the Website 214

Appendix II: Correspondences between log-millisecond (log-ms) and millisecond (ms) pause durations 215

Notes 216

References 227

Index 243

List of Figures

2.1 Southerners TALK slow 15

3.1 Four presentations available in SLAAP of the same transcript data 41

3.2 Praat TextGrid for the transcript shown in Figure 3.1 42

3.3 SLAAP screenshot showing a transcript line with phonetic data 44

3.4 SLAAP screenshot of transcript summary list for Robeson County 46

3.5 Excerpt of SLAAP screenshot showing summary statistics for the transcript for media file ptx0120b 46

3.6 Screenshot of SLAAP's speech rate analysis tool 47

3.7 Screenshot of SLAAP's silent pause analysis tool 48

4.1 Praat Editor window showing one of the reading passages 57

4.2 Considering rate of speech as a slope line 61

4.3 Syllable count and articulation rate measurement distributions 63

4.4 Pause duration measurement distributions (ms and log-ms) 64

4.5 Graphicalizations of the beginning of six reading passages 67

4.6 Articulation rates for reading passage data by utterance and by talker 68

4.7 Articulation rates by talker and speaking rates by talker 69

4.8 Articulation rates by talker and median syllables per utterance by talker 70

4.9 Articulation rates by utterance time for each talker 72

4.10 Effects in the mixed-effect model for reading passage articulation rates 77

4.11 Pause *N*s and pause durations by talker 79

5.1 All speakers plotted by age 88

5.2 Mean utterance articulation rates by main factors 91

5.3 Effects in the mixed-effect model for articulation rates 94

5.4 Mean pause durations by main factors 99

5.5 Effects in the mixed-effect regression model for pause durations 100

5.6 Mean speaker (median) articulation rates by main factors 103

5.7 Median articulation rates by median utterance lengths (MEDSYLS) and median pause durations (MEDPAUSEDUR) 104

5.8 Effects in the fixed-effect regression model for articulation rates 106

5.9 Median syllables per utterance for the speakers 108

5.10 Mean speaker (median) pause durations by main factors 110

5.11 Median pause durations by median utterance lengths (MEDSYLS) and median articulation rates (MEDARTRATE) 111

5.12 Median pause durations by number of pauses per 100 words (PP100WDS) 112

5.13 Effects in the fixed-effect regression model for pause durations 114

6.1 Changes in median articulation rates as sample size is decreased 124

6.2 Comparison of model results for four sample sizes 129

6.3 Pause distributions 133

6.4 Stepwise comparison of minimum threshold increases on pause duration modeling 134

6.5 Comparison of pause model results for different threshold values 137

6.6 Praat Editor window showing an IP-coded transcript for data analysis 140

6.7 Correlation between rates from the main analysis of Chapter 5 and the IP-based analysis 141

6.8 Syllable distribution in all IPs 143

6.9 Effects in the mixed-effect regression model for IP-level articulation rates 146

6.10 Correlation coefficients for the relationship between FF and PFF articulation rates and overall utterance rates 147

6.11 Mean pause durations for subset data by extended factors 150

6.12 Effects in the mixed-effect model for the pause duration subset data 155

7.1 Effect of number of participants on articulation rate and pause duration 161

7.2 Effect of interviewer and interviewee sex on articulation rate 163

7.3 Effect of interviewer and interviewee sex on pause duration 164

7.4 Effects of different/same ethnicity of interviewers and interviewees on articulation rate and pause duration 165

7.5 Speech rate and pause duration medians for EH and her interviewers 169

7.6 Distributions of speech rate and pause duration data for DC females 172

7.7 Speech rate and pause duration correlation for DC interviewees 173

7.8 Pause duration and speech rate comparison for C and her interviewees 175

7.9 Distributions of DC speech rate and pause data, including C 175

8.1 Example of a Henderson graph for an interview dyad 192

8.2 SLAAP screenshot of a Henderson graph 195

8.3 Mean slopes for DC speakers 198

8.4 Effect from mixed-effect model for DC (ing) 205

List of Tables

4.1 Reading passage summary data 65

4.2 Best mixed-effect model for (trimmed) reading passage articulation rate data 75

5.1 Speaker demographics 86

5.2 Best mixed-effect model for (trimmed) utterance-level articulation rates 93

5.3 Mixed-effect (M-E) and analogous fixed-effect (F-E) model fixed-effect coefficients 97

5.4 Best mixed-effect model for (trimmed) pause-level pause durations 99

5.5 Best fixed-effect model for speaker-level articulation rate 105

5.6 Best fixed-effect model for speaker-level pause durations 113

6.1 Speaker demographics for the speakers who contribute more than 100 utterances 126

6.2 Mixed-effect model for the 80 speakers with the most data 127

6.3 Mixed-effect models for the full data, 80, 40, and 20 tokens sampled from each of the 80 speakers 128

6.4 Mixed-effect models for full data and three different threshold levels 135

6.5 IP-level mixed-effect model for Texas articulation rates 145

6.6 Proportion of data and *N*s for region for main data and subset 151

6.7 Initial mixed-effect model for (trimmed) subset pause duration data 153

6.8 Best mixed-effect model for (trimmed) subset pause duration data 154

7.1	Minor and nonsignificant differences between subset and main data	160
7.2	Best mixed-effect model for (trimmed) utterance-level articulation rates after interlocutor factors added	166
7.3	Interviewer information and data summary for EH	168
7.4	Median pause durations and speech rates for DC females	171
7.5	Median pause duration and speech rate for DC interviewees and interviewer	174
8.1	Some Henderson graph-based variables	196
8.2	Slope summary for DC speakers	197
8.3	Basic mixed-effects regression model for DC (ing) data	202
8.4	Full mixed-effects regression model for DC (ing) data	203

Acknowledgments

This project would not have been possible without the work and contributions of very many people, surely more than I can properly acknowledge here. On the one hand, this book is about speech rate and pause and their analysis through a fusion of approaches that I label, as in the book's title, "corpus sociophonetics." On the other hand, the book is about what we – language researchers – can do when we more generally aggregate and "recycle" audio data, recordings of speech that were collected for different purposes than the project at hand. As such, it takes advantage of thousands of hours of work by a large and diverse group of people, from the "master minds" of the original sociolinguistic field projects which produced the interview recordings, to the individual fieldworkers who collected the interviews, to my more recent collaborators who have digitized, organized, data-entered, and helped to transcribe these recordings over the course of the history of the Sociolinguistic Archive and Analysis Project (SLAAP). The best I can think to do here is to thank all of the past and present (and future) members of the North Carolina Language and Life Project (NCLLP), for all of their hard work in the field, in the office, and in the lab, and for their steadfast support of the development of SLAAP. I have built the SLAAP software and the archive framework, but there is no doubt that the archive would be empty without their work. I do thank explicitly those past and present members of the NCLLP with whom I have worked most closely and to whom I feel most indebted: Jeannine Carpenter, Phillip Carter, Erin Callahan-Price, Danica Cullinan, Charlie Farrington, Drew Grimes, Kirk Hazen, Sarah Hilliard, Mary Kohn, Christine Mallinson, Jeffrey Reaser, Ryan Rowe, Natalie Schilling, James Sellers, and Leah White. Erik Thomas and Walt Wolfram have provided tireless leadership during the development and maintenance of SLAAP and, as you will see, I thank them multiple times here. For instance, I thank Walt a second time for being such an inspirational and gracious mentor and for creating the NCLLP in the first place.

Just as the collection of audio recordings I examine here is the product of a massive, joint effort, the fine-grained time-aligned transcripts that form that backbone of my studies are the result of many people's hard work. Many members of the NCLLP, students at North Carolina State University, Duke University, and the University of Oregon – more

people that I can possibly thank here – have contributed to the transcription collection in the archive. Every transcript used here, however, was finalized (i.e. was hand-checked and added to SLAAP) by myself and/or Erik Thomas, who receives his second thanks here for his diligence and selfless commitment to advancing SLAAP. Later in this book, at places of relevance, I thank individual and additional colleagues for more specific collaboration and contributions.

This book and the studies it reports originated in my doctoral dissertation (Kendall 2009) at Duke University. I continue to be grateful to my dissertation committee – Walt Wolfram, Erik Thomas, Ron Butters, and Agnes Bolonyai – for their guidance and mentorship in that period and for their continued friendship, support, and insight as this project has continued over the past few years.

Many people have given me advice on this project over the years – from audiences at conference papers and other presentations to readers of various drafts of this manuscript. Most recently, I am grateful to Erik Thomas, Valerie Fridland, Vsevolod Kapatsinski, two anonymous reviewers, and Olivia Middleton, my editor at Palgrave Macmillan, for comments and suggestions on parts of the book's manuscript. I also thank Gerard Van Herk, Dominic Watt, and Carmen Llamas for many rewarding conversations about the use of Henderson graphs for investigating the realization of sociolinguistic variables (the pursuit of Chapter 8). Charlotte Vaughn has been a constant sounding board and source of good advice throughout this project. I cannot thank her enough. It goes without saying that any errors in this work are my own.

I have received intellectual and financial support from numerous groups over the course of this project. I am indebted to Ann Bradlow and the Speech Communication Research Group at Northwestern University for support during the 2009–10 academic year and to Frans Gregersen and his colleagues, in particular Nicolai Pharao, at the Danish National Research Foundation Centre for Language Change in Real Time (LANCHART) for a visiting research appointment in the fall of 2011. The North Carolina State University Libraries, and their director, Vice Provost Susan Nutter, have been a model of an empowering and supportive academic library. Many other people at the Libraries, including specifically Kristin Antelman, Carolyn Argentati, Amanda French, Greg Raschke, Wesley Thibodeax, and Maurice York, have been integral in developing and maintaining SLAAP as have other members of the Libraries' Digital Libraries Initiative. While this book is not the place to articulate this in full, the relationship between the NCLLP and the university Libraries seems to me a model of library–researcher partnerships.

The data in SLAAP and analyzed in Chapters 5 through 8 were collected in projects funded by the National Science Foundation (NSF) grants BCS-0843865, BCS-0236838, BCS-9910224, SBR-9319577, and SBR-9616331 to Walt Wolfram, grant BCS-0542139 to Walt Wolfram and Erik Thomas, and grant BCS-0213941 to Erik Thomas, at North Carolina State University. The reading passage data examined in Chapter 4 were collected with funding to Valerie Fridland, at the University of Nevada, Reno, from NSF grant BCS-0518264 and to myself, at the University of Oregon, from NSF grant BCS-1122950. I thank the NSF for their continued support of the advancement of linguistic science.

TYLER KENDALL

Part I
Speech Rate, Pause, and Corpus Sociophonetics

1
Looking Forward

1.1 Introduction

This book is about speech timing and, more specifically, about variation in the temporal features of speech rate and silent pause in spoken American English, as viewed from a quantitative sociolinguistic, and to a lesser degree psycholinguistic, perspective. Although it is a book explicitly about the sociolinguistics of speech rate and pause, it is also a book more broadly about corpus-based methodologies and about conducting large-scale sociophonetic research. Throughout this book, I attempt to give as complete an overview of the corpus-based methods and statistical maneuvers I employ as I can. As such, I also provide many resources connected to this book on its website – http://ncslaap.lib.ncsu.edu/speechrateandpause/ – including electronic versions of some data files and tools for, for example, counting syllables in English language orthographic transcripts. It is – of course! – my hope that this project contributes towards our substantive understanding of patterns of speech timing in human language, but I also hope that readers find it useful as a guide to doing large-scale, quantitative sociophonetic research.

In many ways, this book is also about recycling older sociolinguistic recordings and mining them for new phenomena and for the exploration of new questions. It follows from a thread of my research on corpora and data in sociolinguistics (Kendall 2007a, 2008a, 2009, 2011, forthcoming a, b). In particular, while I do not intend this book to be a revision of my PhD dissertation (Kendall 2009), it picks up from that work. There, I discussed in detail the Sociolinguistic Archive and Analysis Project (SLAAP; a web-based sociolinguistic data management system I built at North Carolina State University)[1] and meta-theoretical questions about

data, their treatment, and representation in sociolinguistics, and then turned to a preliminary examination of speech rate and pause as an exploration of how the approach to data implemented in SLAAP made such investigations possible. After several years of continuing to think about and study variation in speech timing, this book presents a much more focused and complete treatment of the sociolinguistics of speech rate and pause. Other than a brief overview of the relevant background in Chapter 3, I leave the larger meta-theoretical discussions of data and data management to the other outlets listed above.

With the goal of examining speech timing in depth, as indicated by the book's title, I limit my focus to patterns of SPEECH RATE and SILENT PAUSE in human language. Other temporal factors, such as segmental durations and speech rhythm, are of interest – and hopefully illuminated upon by the specific foci of this project – but for the sake of maximizing depth in my coverage, I do not pursue them in any explicit way. Pauses, both silent and filled (e.g. *uh*, *um*), are extremely interesting from a number of perspectives, but I will maintain a view on pause from a temporal perspective, focusing more on pause durations than on other potential areas of research, such as the clausal location of pauses, the frequencies of pauses, or the semantics of filled pauses.[2] (Although I will from time to time touch on these subjects, for example, by considering the role of pause location and pause type on silent pause duration in §6.5.)

While this book focuses closely on speech rate and pause, it is also a book more generally about where we find socially differentiated linguistic behavior, the STRUCTURED HETEROGENEITY of Weinreich, Labov, and Herzog (1968). It is about what variation in language can be accounted for by readily enumerable linguistic and social factors. It is about how much of the apparent messiness of variable temporal features – specifically the rate at which speech is uttered and the length of a mid-sentence pause – can be modeled thanks to the burgeoning quantitative and statistical techniques available to the social sciences of the early twenty-first century. At the same time, this book is about what cannot be modeled in this way. It is about what variation is unaccounted for in a large-scale corpus-based analysis, and, better yet, what light we can shed on the processes at work in language production from the unaccountable bits.

Importantly, it is a book about doing CORPUS SOCIOPHONETICS. In these pages, I ask what new things we can learn from treating the large collection of sociolinguistic recordings housed in the SLAAP archive, which were originally collected for various, unrelated sociolinguistic projects, as a coherent sociolinguistic corpus. And I ask the broad question of what the large amount of data obtained through corpus-based analysis (here, ~30,000 measurements of each of the dependent variables) gets us that a smaller dataset does not. Do we learn more from 1000 tokens of a variable from each speaker than we do from, say, five, or from a single estimation of each speaker's general tendency?

Over the past half-century, sociolinguistic research has collected a huge amount of naturalistic speech data. Typically,[3] these data have been used by their collectors to investigate specific research questions and then, after active use over the course of some period of time, the data are put aside and new data, from new communities and research sites and with an eye to new questions, are collected. In recent years, there has begun to be a change in the way that sociolinguistic data are collected and conceived across the discipline. Partly, this is a result of an increasing ability for and interest in conducting REAL-TIME research on language change – that is, to examine comparable data from multiple points in time to examine language change (see Bailey 2002, Sankoff and Blondeau 2007, Gregersen 2009). But, partly, this is a more general result of a reconsideration of sociolinguistic data as corpora (cf. Beal, Corrigan, and Moisl 2007a, b, Kendall 2008a, 2011). Along with the growing sense that sociolinguistic recordings are useful in the long term is a growing sense that they ought to be more "public" than in the past. As Gerard Van Herk and I wrote: "The previous, dominant model of considering sociolinguistic data as too valuable to 'part with' or to share appears to be giving way to a model where sociolinguistic data is considered to be too valuable not to share" (Kendall and Van Herk 2011: 3).

1.2 Disciplinarity and intersections

The past 50 years of sociolinguistic work have also demonstrated the great extent to which systematic variability is a pervasive and integral part of human language. As Weinreich et al. (1968) wrote, a language without variability is both nonfunctional and inconceivable. Variability in form, in structure, and in meaning allows human language its range of expressiveness, its ability "to do things" (e.g. Austin 1962[1975], Searle 1969), and, finally, its ability to change. Variation in language is the explicit focus of research in many areas of sociolinguistics, especially the VARIATIONIST tradition associated with the work of William Labov (e.g. 1966[2006], 1972) and the growing field of SOCIOPHONETICS (cf. Thomas 2002a, 2011a, Foulkes and Docherty 2006). This book grows out of these traditions, but it also seeks to be about something more. In these pages I attempt to connect work in sociolinguistics to other research paradigms

in other areas of language study, in particular within psycholinguistics and social psychology. As we will see in Chapter 2, pauses, and speech timing more generally, have been most actively and productively studied by psycholinguists and social psychologists. Examining these features from a sociolinguistic perspective, but remaining sensitive to the many psycholinguistic findings about them, can aid in our fuller understanding of the nature and function of language variation.

In fact, interest in language variation and, particularly, in how social factors relate to this variability, has grown outside of sociolinguistics in recent years. For instance, work on the psychology of language and within psycholinguistics has often focused on variable features and what that variability means, but most often in terms of what variability shows about speech production on the one hand and how listeners overcome variability as a "problem" for speech perception on the other. Quite recently some of this work has begun attending to the role of subjects' dialect and personal backgrounds more directly. In a 2009 paper published in the *Journal of Memory and Language*, Meghan Sumner and Arthur Samuel examined the perceptual processing of productively /ɪ/-ful[4] and /ɪ/-less New Yorkers and /ɪ/-ful non-New Yorkers and found significant differences both between non-New Yorkers and New Yorkers and between the two New York groups, despite both of the New York groups receiving similar daily exposures to the same /ɪ/-less variants. Instead of stopping there, Sumner and Samuel went on to consider what this may mean for an understanding of "dialect," which despite being widely acknowledged as problematic to define has always been understood (implicitly at least, if not explicitly) as a configuration of *productive* features of a speaker's or group of speakers' language. Sumner and Samuel's results appear to indicate differences in the underlying representations of the forms for these speakers, and the authors suggest that dialects should be considered (or even defined) not only in terms of speakers' productions, but also in terms of their perceptions and representations. They further offer that these three "aspects of a dialect may differ within an individual, just as they differ between individuals" (Sumner and Samuel 2009: 500). Other recent research (e.g. Strand and Johnson 1996, Evans and Iverson 2004, 2007, Hay, Warren, and Drager 2006, 2010, Kendall and Fridland 2012, Fridland and Kendall 2012) has examined the role of social factors on the perception of linguistic forms, but I mention the Sumner and Samuel work because it makes explicit a need for such work, and for sociolinguistic work generally, to consider more deeply its underlying assumptions about "what it means to *have* a dialect" (Sumner and Samuel 2009: 500) in the first place.

Nonetheless, there are of course major differences between sociolinguistics and psycholinguistics. Psycholinguistic research is most often undertaken in the laboratory, in highly controlled settings, while sociolinguistic research is most often undertaken in the field in settings and ways that might maximize the naturalness of the spoken language, that is, that minimize the OBSERVERS' DILEMMA (cf. Labov 1972, Milroy 1987) rather than control the possible sources of variation. It is also true that the main research questions of sociolinguistics and psycholinguistics differ greatly. Yet, I believe it is fair to say that each of these fields studies variation and is interested in what that variation *means*. For sociolinguists, interest is often in variation because it yields insight into the extralinguistic, social factors in language in use, and, for scholars who follow Labov's variationist paradigm, in that understanding variation is central to understanding language change. For psycholinguistics, variation is often useful as a window into the processes of language production and a source of potential difficulties in language comprehension and processing. Variation pervades both of these fields and both have yielded great insight into the causes and meanings of that variation. Yet, for most of their histories, research in these fields has operated independently. To make an observation that is surely overly simplistic: sociolinguists publish in sociolinguistic journals and psycholinguists publish in psycholinguistic journals. There is just too much to read (and moreover to do) for us to follow everything of interest. Yet, to understand variation and its role in human language more fully greater collaboration is needed across these disciplines. Perhaps the time is right to pursue a more collaborative SOCIAL PSYCHOLINGUISTICS?

But a label is just a label, and, while I think this label invokes some ideas worth considering, my goal is not to dwell on terminology in these pages. Further, this book is surely not the first place to consider such a thing as a social psycholinguistics (though the collocation is surprisingly rare).[5] As I mentioned earlier, psycholinguists and laboratory phoneticians have recently begun to pay closer attention to the literature on socially differentiated language variation (such as the work by Sumner and Samuel). The burgeoning field of sociophonetics (cf. Thomas 2011a, Di Paolo and Yaeger-Dror 2011), with its instrumental and often experimental methods, bridges some traditional gaps between these research disciplines. (Readers are referred to Thomas 2011b for a recent review of work relating sociolinguistic variation to cognition.)

So, while I write this book primarily as a sociolinguist, I see the boundaries of these two approaches – sociolinguistics and psycholinguistics – as overlapping, and ultimately, almost nonexistent. Where

do social factors disappear or become irrelevant? Where do cognitive factors cease to impact language production and perception? I approach the questions of this book from the view that separating these two sets of factors within a thorough study of actual conversational speech is about as possible as imagining a language without variability.

As I wrote above, a major disciplinary difference between the importance of variation to sociolinguistics and psycholinguistics is how that variation informs our understandings of language and our theoretical perspectives on language. A second major difference has traditionally been in methodology. The field-based studies of sociolinguistics are a kind of corpus-based linguistics, with the fieldwork generating richly contextualized corpora of natural speech data.[6] Psycholinguistics, on the other hand, has traditionally used experimentation to gather its data and test its hypotheses. Increasingly, however, these methodological differences are blurring and numerous sociolinguists have taken to lab-based, experimental methods (cf. e.g. Campbell-Kibler 2005, 2007, 2010, Hay, Drager, and Warren 2009, Drager 2010; see Thomas 2002b for a thorough and historical review). Psycholinguists have also increasingly incorporated (most often standardized) corpora and corpus analyses into their research projects (e.g. Clark and Fox Tree 2002, Bell et al. 2003, Kapatsinski 2010, just to list a few). Ultimately, I believe that both of these approaches to empirical linguistic analysis are necessary to better understand language variation, change, and processing. Nonetheless, in this book, I limit my focus to corpus-based examinations. Several of my suggestions and findings in later chapters point to the need for further experimental testing and doing so would surely strengthen the findings of this research. However, for space, time, and focus, I maintain a strictly corpus-based view here, with the aim of exploring just what we can learn from such an approach.

1.3 Why exactly speech rate and pause?

It is worth in this first chapter to ask why we might want to study speech rate and pause rather than some other features. Especially as a linguist and a sociolinguist, why should I (or you for that matter) be interested in these features, beyond the fact that they are amenable to large-scale corpus-based analysis? The answer, I believe, is as follows.

Rate of speech and pause are ubiquitous features of human language. Every utterance by every speaker of every language (even sign languages) can be characterized as having a particular rate of production and by being in relation to some intervals of silence. Further, 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. For some of the source data examined in this book, as much as 35 percent of the transcribed recording is in fact silence on the part of the participants! (Admittedly, these high numbers are from particularly reticent participants and figures of about 15 percent are more typical.) By looking closely at these omnipresent phenomena we can gain insight into larger patterns of variation, and variation in less common features.

A related question then would be why do I *only* examine speech rate and pause. Other temporal features – e.g. segment durations and speech rhythm – are also relevant here and would also be usefully examined in the context of a large-scale corpus sociophonetic analysis. The answer here is two-part. Practically speaking, I limit my focus to these two features for sake of time and space. I seek to be comprehensive in this monograph in my description of their study and adding more features, even related ones, would make this project too unwieldy. More importantly, I focus on these two features specifically because of their joint role in the way that listeners hear speech rates. As we will explore, much evidence has pointed to the role of pause durations in the perception of rate differences and it seems to me that a study of variation in speech rate would be incomplete without a close attention to variation in pausing as well.

I am not the only one to take a recent sociolinguistic interest in temporal features in speech – variation in speech timing appears to be an area of growing interest in linguistic research. As Chapter 2 will address, some recent work has examined pause and speech rate from a sociolinguistic perspective. Other features, like speech rhythm (something I do not examine in this book), have also been the focus of some recent sociolinguistic research (e.g. Thomas and Carter 2006, White and Mattys 2007). Understanding the naturally occurring variation in these features is important at a number of levels. From linguistic and sociolinguistic theoretical perspectives, establishing whether these features correlate with social attributes of speakers has ramifications on theories of grammar and on the social influence on language. For example, at what levels of fine phonetic detail do we find patterned variation? Where does this patterning break down into the noise of so-called FREE VARIATION? Is there such a thing as free variation? From a purely empirical perspective, opening up all of this silence and temporal data to analysis creates new opportunities for phonetic and computational analysis. Finally, as I will consider at length in Chapter 8, once we have

a grasp on the social and cognitive factors that influence speech rates and pause durations, we can then shift our attention *back* to the utility of these features as potential predictors behind the realization of phonological and morphosyntactic variables. And it is here, perhaps, that an attention to pause and speech rate can most fully benefit the quest to understand the principles and processes underlying language variation and change.

1.4 Overview of the monograph

The remaining two chapters of Part I provide overviews of areas of linguistic research and background related and relevant to the present project. In Chapter 2, I consider what we know about pause and speech rate and attempt to bring together findings from the quite disparate traditions that have approached these questions. The previous research on pause and speech rate is used to develop a general understanding of the source and meaning of variability in these features. It also lets us develop some expectations for the empirical analyses of Part II. In Chapter 3, I back up to explain the origins of this project and its foundations in my work on archiving and managing sociolinguistic data. In that chapter, I also explain the underlying transcript model that forms the basic data from which speech rates and pauses are measured and the tools that I use to extract those measurements.

Part II represents the bulk of the book and presents a number of empirical, original studies on speech rate and pause. I begin, in Chapter 4, by examining speech timing in a small multiregional corpus of read speech recordings in order to discuss the general framework of analysis and basic methodologies of the study. This small analysis finds some social differentiation in the data, especially for speech rate, but I ultimately argue that read speech is far from ideal for studying patterns in speech timing. This motivates Chapter 5, the largest (datawise) study in the book. Here I consider speech rates and pause durations from about 30,000 measurements each, taken from the English speech of 159 individuals from areas in the United States (Ohio, Texas, Washington, DC, and primarily North Carolina). In this examination, I show that speech rate patterns quite strongly according to speakers' basic social factors (region, ethnicity, sex, age). Pause variation, on the other hand, while exhibiting some social correlations, does not pattern strongly with social factors.

In Chapters 6 and 7, I continue the corpus-based investigations of speech rate and pause. Chapter 6 focuses on four additional corpus-based questions as a further development of the line of enquiry of Chapter 5: how

many measurements are needed for stable patterns; how long is a pause; a comparison of rate data coded at the Intonational Phrase rather than phonetic utterance level; and a second attempt to account for the pause duration variability by considering additional potential factors. Chapter 7 turns its attention to other sociolinguistic kinds of factors, in particular to within-speaker variation and the influence of interlocutors on speakers' rates and pauses. Throughout these studies we continue to obtain robust patterns for speech rate and a noisier picture for the pause data.

In Part III, I attempt to take advantage of the accumulated knowledge about speech rate and pause to advance the sociolinguistic study of language variation. While the empirical studies of Part II indicate that speech rates are systematic across and within speakers, they also indicate that pauses are not. However, patterns in pausing – as will be discussed in Chapter 2 – have a long-studied relationship with cognitive factors, and this, I propose, allows us to use pause variation as a way to better understand the realizations of other, more commonly studied, sociolinguistic variables. Thus, in Chapter 8, I reconsider the notion of CHANNEL CUES TO ATTENTION TO SPEECH (Labov 1966[2006], 1972) in terms of our larger knowledge of pauses. I revisit a technique, the HENDERSON GRAPH, from an early line of psycholinguistic pause research (Henderson, Goldman-Eisler, and Skarbek 1966) to examine sociolinguistic variation, and I show that this method captures a relationship between speaker hesitancy (measured in terms of pause-to-talk time) and the realization of variable (ing), the alternation of *-ing* and *-in'* in words like *talking* and *something*. From this, I propose a framework for future research, which might allow us to assess the cognitive status of various kinds of sociolinguistic variables (Thomas 2011b). Finally, I end the book with a short assessment of this whole endeavor and a discussion of where this might lead the study of language variation.

2
What We Know about Speech Rate and Pause

2.1 Introduction

When I first set out on this project, I envisioned including a truly comprehensive review of all of the work that had been done to date on speech rate and pause in human language. Within the field of sociolinguistics, this would be a short task, as interest in temporal features has mostly been little and sporadic. In general, many dominant views of language, such as those in the Saussurian and Chomskian traditions, have traditionally placed the study of pause and speech rate outside of the realm of linguistics proper. From a structural or generativist perspective, pause and speech rate are so clearly components of LINGUISTIC PERFORMANCE rather than LINGUISTIC COMPETENCE that they have had no place in formal approaches to linguistics. However, even within variationist and sociolinguistic work, areas which have made great headway by studying linguistic performance, there has been a long history of considering speech rate and pause as nonlinguistic and, frankly, not relevant.[1] In a widely read sociolinguistic handbook chapter, for instance, Ronald Macaulay wrote:

> 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)

Despite this view that pauses (and, perhaps, by extension speech rates) cannot be studied in linguistic terms, we find a very different perspective when we turn our attention to the psycholinguistic and psychological literature. There, the study of sequential temporal patterns, and pauses in particular, have been a major area of focus, at least among some specific groups of scholars. For instance, Frieda Goldman-Eisler, whose work I will return to at length, published a 1968 book titled simply *Psycholinguistics: Experiments in Spontaneous Speech*, which was more than anything else a review of about a decade's worth of her experimental work on pauses in spontaneous speech. Researchers, such as Sabine Kowal, Daniel O'Connell, and Stanley Feldstein, have devoted large parts of their careers to the understanding of pauses and other speech-timing features. Speech timing has been a central component of the work in social psychology on interpersonal interaction and Howard Giles' development of COMMUNICATION ACCOMMODATION THEORY (CAT; an area that has directly influenced the direction of modern sociolinguistics). Meanwhile, work on first and second language acquisition has attended to pause and temporal patterns as indications of progress during language acquisition and as something to acquire in their own right (e.g. Clark 2009, Redford in press). Computer scientists and computational linguists, like Julia Hirschberg (e.g. Edlund, Heldner, and Hirschberg 2009, cf. Zellner 1994), have also attacked the problem of understanding patterns of pauses and speech rates in their quests to improve speech recognition and to develop naturalistic speech synthesis. As I expanded away from the purely sociolinguistic in my quest to understand what has been learned about speech rate and pause patterns, it became clear that a truly thorough review of the literature – that an accounting of everything we know when it comes to these features – is an impossibility. And, actually, one of the striking observations that comes from a broad survey of the broad literature is just how many patterns have been found for pauses and speech rates and the wide range of phenomena which have been claimed to relate to these features. Reviewing this wide literature leaves one with the sense that pauses and articulation rates pattern with everything!

While not claiming to cover the topic in its entirety, in this chapter I assess the state of our knowledge about speech rate and silent pause, and their variability. I focus primarily on the sociolinguistically relevant findings in the literature, and I use this space to better spell out the main questions that are to be asked in a large-scale, corpus-based sociophonetic study of these features. I begin by reviewing evidence – both scholarly and folk-based – that speech rate and pause are socially meaningful to speakers and hearers.

2.2 Attitudes towards and the perception of speech rate and pause

While temporal factors of speech, like speech rate and pause duration, have not been of great interest in the history of sociolinguistics, there is plenty of evidence that these aspects of speech timing influence popular conceptions of dialect differences and listeners' social judgments of others. That is, listeners perceive, and expect, differences in speech timing based on a number of social factors, and these kinds of temporal factors are a central part of what Deborah Tannen has termed CONVERSATIONAL STYLE (Tannen 1984[2005], 1985, 2000), the discourse-level differences that mark, for instance, New York City Jewish English (e.g. Tannen 1985) with its heavy use of overlapping turns and so forth as markedly different than, say, varieties of Native American English, which are often described as valuing silence and long pauses between turns (Philips 1976).

Most famously, in the US, is the popular myth that Southerners talk more slowly than non-SOUTHERNERS. In fact, the common term for a Southern accent – a SOUTHERN DRAWL – is by its very definition a portrayal of Southern speech as not only accented, but slowed or even affectedly slowed.[2] The classic film about language variation in the US, *American Tongues*, features a Texan columnist, Molly Ivins, who provides a nice example of the markedness of Southern speech and its association with the slow-talking stereotype:

> There's a lot more prejudice against a Southern accent than there is against any other kind. That is- and I think it troubled Jimmy Carter considerably, because in the Northern mind a southern accent equals both ignorance and racism and you'll see that stereotype reinforced in zillions of old movies. You take all those old movies, around World War II era. I don't know how many zillions there were but the classic World War II movie consists of an "All-American" clean-cut hero who was from somewhere in the middle-west. Usually a farm kid from Kansas, who's blond and he's always got one wise-cracking buddy from New York and then *there's always some just dumb, slow-talking Southerner who's the butt of all the jokes* in the military movie. And that's a- that's a stock character in American movies and it really has reinforced the prejudice against the southern accent. (Alvarez and Kolker 1988: 30.05, emphasis added)[3]

The view provided in Figure 2.1, from a bank advertisement in a Southern newspaper, shows that the slow quality of Southern American English is

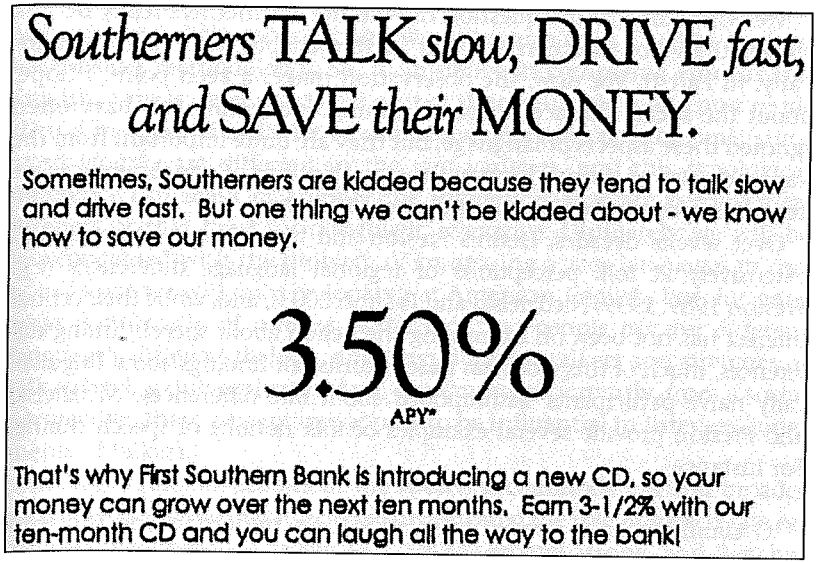


Figure 2.1 Southerners TALK slow. *Times Daily Newspaper*, April 27, 2005

not only a trope even within the South but also something that can be reclaimed – or at least used for humor and marketing. (A Google search for "Southerners talk slow" appears to retrieve as many positive associations with slow-talking Southern English as negative ones.)

In recent work, Tyler Schnoebelen (2009, 2010) has investigated speech tempo from the perspective of INDEXICAL FIELDS (Silverstein 2003, Eckert 2008) and nicely demonstrated the richly interwoven web of meanings which are associated with slow- and fast-talking speech. For instance, developing visual, indexical fields for these two speech types based on the ways that they are described in several corpora, Schnoebelen (2009) demonstrates that slow speech is associated with "Southern" talkers and "surfers," with "introverts," with "incompetent" speakers and "liars" but also with "thoughtful" and "articulate" talkers and with "doctors." Silverstein's notion of INDEXICALITY, and Schnoebelen's work on tempo from this perspective, nicely allow for the exploration, and larger coexistence, of these at first seemingly contradictory meanings. The specific ways in which slow (or fast) speech is perceived at a given moment are resultant from and a part of the speaker's larger stylistic package, his or her "conversational style" (going back to Tannen), and the larger discourse and social context.

We will return to the question of whether Southerners really do talk slower than non-Southerners later in this chapter and again, empirically, in Part II. For now, the observation makes a keen point, I hope, about the social salience of speech timing. (Socio)linguists have often ignored these aspects of language, but they are quite important from the vantage point of normal human listeners. Notions of rate differences are central to popular beliefs about dialect difference.

Over several decades, Dennis Preston and his colleagues have looked extensively at folk perceptions of regional language differences (e.g. Preston 1989, 1999, Niedzielski and Preston 2003), and, while their central interest has not been on examining the beliefs about speech timing differences, much of this work has raised consistent findings about linguistically naïve participants' assumptions about rate differences. Niedzielski and Preston provide several examples of folk notions of speech timing. For instance,

> G claims to have needed translation help understanding Southern when he was in the [military] service, although a characteristic of the variety was its speed:
>
> G: uh- I was stationed in- in- in Georgia for a while, stationed in Fort Monmouth New Jersey, (.hhh) an:d I had to look at two of my buddies to sometimes figure out what somebody was saying. (.hhh) When they-they talk in a Southern draw, (.hhh) and I would wait for the words to finally come out because they go real: real: slow. (Niedzielski and Preston 2003: 109–10, example edited to remove nonrelevant interrupting speech by interlocutor)

It is also worth noting that these assumptions or intuitions about speech timing differences are not just held by nonlinguists. For instance, in a paper investigating rural vs urban differences in speech timing, Hewlett and Rendall (1998: 63–4) point out that John Wells, the famous British phonetician, claimed "[i]t is perhaps universally true that rural accents tend to be slower in tempo [than urban ones] reflecting the unhurried life of the countryside"[4] (Wells 1982: 11) but later amended that this "universal" is "an impressionistic claim rather than ... a substantiated fact"[5] (Wells 1982: 87).

In his contribution to the *Language Myths* volume edited by Laurie Bauer and Peter Trudgill, Peter Roach (1998) discusses a similar question to that of differences between regional dialects – whether different languages are characterized by different rates of speaking. Overall, Roach indicates that findings of different rates across languages can

be contradictory and may be more a result of different measures than actual varietal differences (a point we will return to in Chapter 4, when we consider methodology). So, for instance in comparing Finnish and English, a measure of words per unit of time will yield a different result than a measure of syllables or sound segments per unit of time, since word lengths are different in the two varieties, and this typological difference can mask or amplify the differences that are perceived by listeners or that have sociolinguistic relevance. Ultimately we will be less concerned with the problem of comparing across languages in this book, since we will only be looking at American English, but, for now, what is primarily of interest is the fact that people *perceive* different languages, different dialects, and even different talkers and stretches of talk as having different rates. And, even more importantly, from a social perspective, these perceptions appear to be influential in listeners' judgments of talkers.

In fact, numerous studies in the speech accommodation and broader social psychological literature have investigated the role of speech rate on listeners' judgments of talkers (e.g. Smith, Brown, Strong, and Rencher 1975, Miller, Maruyama, Beaber, and Valone 1976, Apple, Streeter, and Kraus 1979, Giles and Smith 1979, Brown 1980, Thakerar and Giles 1981, Street and Brady 1982, Street, Brady, and Putnam 1983, Street, Brady, and Lee 1984, Giles, Coupland, Henwood, Harriman, and Coupland 1990, Ray, Ray, and Zahn 1991, Ray and Zahn 1999) and have yielded numerous corroborative and consistent findings. For example, faster speech is typically associated with competence, intelligence, expertise (Smith et al. 1975, Thakerar and Giles 1981, Street and Brady 1982), "social attractiveness" (Street et al. 1983; but for male voices only in Street et al. 1984), and greater persuasiveness (Miller et al. 1976, Apple et al. 1979) over slower speech. However, Giles and Smith (1979), Street and Brady (1982), and others have shown that "speech rate preference regions for socially attractive others [often center] around the *receiver's* typical speech rate level" (Street et al. 1983: 39; emphasis in original), and, notably, Apple et al. (1979) and Smith et al. (1975) found indications of U-shaped patterns, where fastest rates were perceived to be less "truthful" and "benevolent" than rates in the middle. Further, much of this evidence has indicated that listeners are more sensitive to rate differences in making these kinds of judgments than they are to other aspects of accent (like in judging New Zealand English accents against American English accents; Ray and Zahn 1999).

As the above implies, there is evidence that speech rate perception is mediated by listeners' social and communicative expectations.

Street et al. (1983), for instance, found some evidence that listeners were more aware of speech rate differences when told they were listening to talk in an "employment interview" context as compared to a "conversation" context. Siegman and Reynolds (1982) indicate that speech rates are expected or allowed (i.e. interpreted favorably) by listeners to be slower in "highly intimate settings." While much research has shown that faster speech rates are associated with qualities of competence and intelligence and so on, some research has also shown that higher-status talkers are perceived as talking faster than lower-status talkers (Thakerar and Giles 1981, Thakerar, Giles, and Cheshire 1982) regardless of their actual rates. (Might the stereotyped perception of US Southerners as slow talkers have more to do with social valuations than actual speech production?)

Finally, it is also clear, from areas of research like CONVERSATION ANALYSIS (see e.g. Liddicoat 2007) and DISCOURSE ANALYSIS (e.g. Johnstone 2007), that fine-grained timing features like pause and speech rate play crucial roles in meaning-making at the utterance and discourse level. Norma Mendoza-Denton (1995) provides an excellent example of the power of pauses in her paper "Pregnant pauses: Silence and authority in the Anita Hill–Clarence Thomas hearings." This work demonstrates the way that subtle differences in gap length – silence between turn changes in discourse – both reveal the ideologies of participants in the Senate hearings and shape the interpretation of the discourse for observers. In sum, pause and speech rate variation appear to be of far-reaching importance in actual talk and interpersonal interaction.

People may readily talk about "slow" dialects, "fast" talkers, "long" pauses, and so forth, but what exactly in acoustic terms are they really talking about? That is, what are people listening to when they hear rate and pause differences? Goldman-Eisler's important work on pauses (1968) suggested that in judging speech rates listeners may be attending more to the distribution and length of pauses and not (or less so) to the actual rate of speech production. Other research – such as the limited work on the Southern drawl (e.g. Wetzell 2000) – has considered the role that vowel duration or even the spectral dynamics of vowels (monophthongization, diphthongization, and so forth) may play in listeners' percepts of speech rate. So while listeners may be quite sensitive to slow and fast speech, they are not necessarily attending to rate per se. However, despite Goldman-Eisler's strong stance on the role of pauses and my own focus later in this book specifically on pauses and a pause-exclusive measure of ARTICULATION RATE (see §4.4), it remains not all that clear what exactly listeners attend to when they make judgments about

speech rate, or how good they are at discriminating these kinds of differences. There is disagreement in the literature and contradictory findings across studies.

Laver (1994: 542, cited in Roach 1998) argued in his *Principles of Phonetics* that "the analysis of phenomena such as rate is dangerously open to subjective bias ... listeners' judgments rapidly begin to lose objectivity when the utterance concerned comes either from an unfamiliar accent or (even worse) from an unfamiliar language." Yet, there is simultaneously plenty of evidence that listeners are quite good at accurately perceiving rate differences and identifying pause locations. For instance, Vaane (1982) found that both trained and untrained listeners were able to classify the rate of speech for sentences spoken in languages ranging in familiarity (native, familiar, and unfamiliar) with roughly the same degree of accuracy. The social psychological and speech communication literature in testing attitudes towards rate differences has also confirmed this – for example, Ray and Zahn's (1999) study of attitudes towards New Zealand English found that listeners' perception of rate differences corresponded to actual rate differences and Robb, Maclagan, and Chen's (2004) study comparing speech rates between New Zealand English and American English also found that listeners' judgments aligned with the acoustic results.

A fairly long tradition of work has examined the perceptibility of pauses, going back to Goldman-Eisler (1968) and work by her contemporaries (e.g. Martin and Strange 1968). In a series of somewhat more recent projects, Duez (1982, 1985) examined what acoustic and linguistic features correspond to the identification of silent pauses in speech and found that identified pauses correlated with prosodic characteristics of the talk (more than linguistic information, such as syntactic location) and that pause duration was a major correlate of perceivability. Duez (1993) also examined SUBJECTIVE PAUSES, pauses that are perceived but that do not correspond to actual silences in the acoustic signal, and further indicated that prosodic aspects of the surrounding talk (such as lengthened vowel duration) can cue pause perception without actual silence.

To the best of my knowledge, there have not been studies that have attempted to measure the JUST NOTICEABLE DIFFERENCE (JND) – the degree of change necessary for a difference to be perceivable to a listener – for pause durations. Such studies would help to shed light on the range of durational differences in pause realizations that listeners can actually discriminate. However, from the vantage point of speech rate, Quené (2007) reviewed the literature on JND and noted the paucity of studies

relevant to speech communication (most have been about tempo in music) and conducted an experimental study to examine the JND for speech rate. His experiments "provide an estimated JND of 5% of the base tempo of a speech utterance. Tempo variations exceeding this [difference limen] are likely to be noticeable, and relevant in speech communication" (2007: 360).

For our purposes, this discussion is meant to indicate both the extent to which pause and rate variability are important sociolinguistic components of language varieties and individual discourses as well as the difficulties and subjective nature of determining the exact relationship between acoustic cues and these larger perceptions and attitudes. Shedding further light on the perception of these aspects of speech timing is, unfortunately, outside the scope of the corpus-based research I pursue here. Given recent advances in research on sociolinguistic perception (cf. Campbell-Kibler 2010), we can hope that future experimental research will add to our knowledge of how and why listeners hear speech as fast or slow. We now turn our attention to reviewing the realization of pauses (§2.3) and speech rate (§2.4) in further detail.

2.3 Pauses in detail

Frieda Goldman-Eisler, the prominent psycholinguist and pioneer of pause studies, described some of her findings thusly:

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

Her work (e.g. Goldman-Eisler 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, as mentioned in the last section, Goldman-Eisler found that pauses account for much of the variation in perceived speech rate.

Much of the psycholinguistic literature on pauses has followed Goldman-Eisler's lead and considered pause to be an outcome and indicator of processing activity, and her work is paralleled by the findings from other psycholinguists who have pursued questions of speech timing. For example, in a well-known 1959 paper – one of the only studies not by Goldman-Eisler from this earliest period – Maclay and Osgood found that hesitation pauses are more often realized before a semantically heavy unit than at clause boundaries.[6] In general, Goldman-Eisler's various findings appear to have been confirmed numerous times and in numerous ways (e.g. Lay and Paivio 1969, Siegman 1979b, Kircher, Brammer, Levelt, Bartels, and McGuire 2004; see, more generally, Levelt 1989). There has been some disagreement on mostly minor points in the early literature (see Boomer 1970, Rochester 1973), but the main findings from Goldman-Eisler's work – for example that pauses increase with task difficulty – have been quite robust across studies.

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"). 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:

> For example, 10-year-old children pause more frequently when telling stories before an audience of adults than when they are alone, speaking into a microphone (Levin and Silverman 1965). Moreover, differential sensitivity to others seems to affect [silent pause] incidence. Subjects scoring high in an audience sensitivity test pause more frequently when addressing an audience than did low scorers (Reynolds and Paivio 1968) but these differences were not found in the absence of an audience (Lay and Paivio [1969]). Pause frequency remained constant but duration increased when utterances of subjects scoring high in concern for approval (Preston and Gardner [1967]) and extroversion (Ramsay 1968) were compared with the vocalizations of low-scoring subjects. (Rochester 1973: 75)

Sabine Kowal and Daniel O'Connell have a long history of interest in "pausological research" (cf. Kowal and O'Connell 1980). They credit the main hypothesis of this line of research directly to Goldman-Eisler, building on the idea 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 their colleagues, however, expanded the range of interest in pause beyond the primarily psychological focus of Goldman-Eisler's work. For instance, they 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, Sabin, Clemmer, O'Connell, and Kowal 1979).

> 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) and argue that pausing is different for children than for adults (Sabin et al. 1979). Occasionally, O'Connell, Kowal, and colleagues have taken an interest in broader social factors in pause (and speech rate) beyond foci on cross-linguistic comparisons and age-graded, developmental patterns (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 also found some evidence that young urban children in lower socioeconomic situations have longer pauses than their higher socioeconomic peers, 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 for the most part they have not been pursued to any depth in the following decades. One exception is a recent study by Redford (in press) which has followed up on Sabin et al. (1979) and examined differences in pause patterns between kindergarten-age children and adults. Redford found some differences between the child data and adult data – for instance, that children produced significantly higher rates of pauses in ungrammatical locations than adults – but did not find evidence that pausing is a different phenomenon for children than adults.

Beginning in the 1970s, Stanley Feldstein and his colleagues undertook a number of connected projects, examining what they termed "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). Some of this work also examined the influence of other personality characteristics on individuals' pause realizations and found that both an individual's personality characteristics and, to a lesser extent, their interlocutor's personality characteristics impacted their pause durations (Feldstein, Alberti, and BenDebba 1979). For example, Feldstein and colleagues tell us that "persons who are reserved, cold, suspicious, insecure, and tense tend to produce longer pauses" (Feldstein et al. 1979: 85). Importantly, a number of Feldstein's experimental findings support the formation of different impressions by hearers on aspects of pause depending on social attributes of the speakers, such as ethnicity and gender (Feldstein and Crown 1978, discussed in Crown and Feldstein 1985; see also Feldstein et al. 1979, 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, they provide evidence that gender and ethnicity interact with speech timing features in influencing speaker-listeners' perceptions of one another. So, while the earlier work in psycholinguistics (such as by Goldman-Eisler 1958, 1968, Maclay and Osgood 1959, etc.) focused on pause as a cognitive, psycholinguistic phenomenon, this work supports the view that pause also has a social component outside of being the outcome solely of mental processes.

Other researchers have investigated these kinds of pause patterns as well. For instance, Aron W. Siegman (e.g. 1979a) and colleagues, in a number of studies of interviewer–interviewee interactions, especially focusing on the impact of interpersonal attraction on these interactions, found that interviewees exhibited fewer and shorter pauses when their interviewers were more socially attractive. I have reviewed some of the most relevant social psychological work on speech timing in the

discussion of attitudes towards speech in the previous section, but, as in Siegman (1979a) and as illustrated by the discussion of Feldstein's research, some of this same research has also examined the production of pauses and the production and perception components of these projects cannot always be teased apart.

While these studies indicate that social factors may have some role in pause production, especially in terms of accommodation to various kinds of audiences, arguments have been made that these social differences and accommodation-like effects may be nonetheless underlyingly related to aspects of cognitive load. For instance, Cappella (1985) wrote,

> Pauses and switching pauses are basically measures of reaction time in the domain of speech and, hence, are a reasonable set of indicators of cognitive difficulty and load. Siegman (1978, [1979b]) has been making just these arguments. The silent pausing associated with ambiguous questions, general questions, intimate interactions, interactions with unattractive and cold persons, and with difficult and unfamiliar questions need not be explained by differential appeals to anxiety, and interpersonal attraction, but through the parsimonious mechanism of cognitive decision making. Each of the above conditions requires greater monitoring of one's choice of words and, hence, places the actor under greater cognitive load. This decision making takes time resulting in greater pausing. (Cappella 1985: 90–1)

Interest in pause has also come from researchers interested in the structure of discourse. An exciting example is found in Wallace Chafe's work on the *Pear Stories* (Chafe 1980a), in which Chafe used pause to help better understand the unfolding of information flow in discourse. 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 pauses occurring between clusters have a longer mean duration than those within clusters (Chafe 1985).

While social differences in pause realization have not been examined to nearly as great an extent as psychological and task-based factors, the

social component of pause realization 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 Deborah Tannen and Muriel 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 with "The Silent Finn" of Lehtonen and Sajavaara (1985; Sajavaara and Lehtonen 1997). As mentioned earlier, pauses can be viewed as a part of conversational styles (Tannen 1984[2005]), but at the same time pause differences appear to exist at a more macrolinguistic level. Campione and Véronis (2002) quantitatively compared pause duration across five European languages (English, French, German, Italian, and Spanish) by analyzing approximately 6000 pauses in about 5½ hours of recorded speech and 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 corpus linguistic and computational linguistic perspectives, 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 it has been found to be a significant constraint in the realization of some variables. The major example of this is CORONAL STOP DELETION (CSD; often also called T/D DELETION or CONSONANT CLUSTER REDUCTION), where numerous studies (e.g. Guy 1980, Wolfram, Childs, and Torbert 2000) have found a following pause to constrain consonant cluster reduction differently than following consonant or vowel environments. (I return to considering pause as an independent predictor in the realization of sociolinguistic variables in Chapter 8.)

As I mentioned above, sociolinguists have recently become interested in understanding prosodic variation and a few recent efforts have begun investigating questions around pause (and speech rate, which is addressed in the next section). In a 2006 conference paper, I asked whether pause could be considered a SOCIOLINGUISTIC VARIABLE (Wolfram 1993) and found favorable results. Recently, Cynthia Clopper and Rajka Smiljanic (Armstrong, Clopper, and Smiljanic 2008, Clopper and Smiljanic 2011) have investigated regional and sex-based variation in

pause (and speech rate) and asked whether pause duration was a factor in the stereotype that Southerners talk slower than Northerners. Their (2011) comparison between read speech from the Midland region and the South found that there were no differences in pause durations by region or speaker sex (and no differences in speech rate). They did find a significant difference for pause frequency, with Southern males having significantly more pauses than the other subjects (Midland males and females and Southern females). Byrd (1994) examined speaking rate in sentence readings across several regions of the US from the TIMIT database (Garofolo et al. 1993) and found that Southern and South Midland speakers had slower rates than other regions – largely as a result of having more pause time. While the longer pauses (and slower rate) of the South make sense in terms of the above discussions of "slow Southern speech," Byrd also found that New York City had the next slowest rates, though this goes contrary to the common stereotype of fast-talking Northerners (and New Yorkers in particular). It is important to remember that these studies were conducted on read speech, as I will have more to say about this source of data for speech timing analyses in Chapter 4.

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 is missing here, and what motivates the present study, is an explicit investigation of the relationship between pause production and social differentiation at a more nuanced quantitative level than between geographically distant and culturally separate populations. That is, do groups (however socially defined) perform or index their group identity through their pause practices? Further, is the variation found in pause realization – such as that found by Campione and Véronis (2002) in their comparison of European languages – systematically related to social factors within languages? Wolfram (2006: 334) 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." Does variation in pause fall within the realm of *significant language variation*?[8]

2.4 Speech rates in detail

Much of the primary psycholinguistic research on the temporal sequencing of speech (such as the work by Goldman-Eisler, Kowal, and O'Connell) has treated speech rate 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) can lend to language planning and production and Goldman-Eisler's early claim that speech rate (measured via articulation rate) exhibited little variation 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)

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. This possible influence of pauses in the perception of speech rate requires an important clarification of terminology. I have thus far been discussing rate of speech in general terms ("rate" or "rate of speech" or "speech rate"). It is necessary to make a distinction between pause-exclusive measures of rates – termed ARTICULATION RATE – and pause-inclusive measures of rates – SPEAKING RATE. (In Chapter 4, I discuss these measures more thoroughly but for now note that I use the terms "articulation rate" and "speaking rate" throughout this book to refer explicitly to these specific measures; terms like "speech rate" and "rate of speech" are used when the distinction is unnecessary.)

While the position that pauses play a primary role in the perception of rate has been supported by others (e.g. Grosjean 1980b), the notion that they are *primary* has also been contested. Miller, Grosjean, and Lomato (1984) demonstrated that variation in speech rate is significant on its own, even within single speech events, and further argue that speech rate variation was significantly underappreciated in the earlier work of scholars like Goldman-Eisler (and Grosjean 1980b). Much other work has also supported the notion that speech rates are in actuality more variable than indicated by Goldman-Eisler. Siegman (e.g. 1979a: 101–2), for instance, found that articulation rates modulated with task manipulations, a finding clearly in line with the kind of variability

Goldman-Eisler was interested in, but only found for pauses. Some other recent considerations have also been critical of Goldman-Eisler's findings about rate and have taken the position that her findings are mitigated by methodological issues (cf. Dankovičová 2001).

So, despite Goldman-Eisler's claim of rate being mostly invariant at the speaker level, later scholars have focused on intra-speaker speech rate variation to some greater extent. Deese (1984) 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. Most researchers have found the opposite, however, that the last few words of an utterance are in fact the slowest. Dankovičová (2001), Yuan, Liberman, and Cieri (2006), Quené (2008), and Kendall and Thomas (2010) all find strong evidence that rates slow at the end of utterances through PHRASE-FINAL LENGTHENING (cf. Beckman and Edwards 1990, Turk and Shattuck-Hufnagel 2007).

Dankovičová's (2001) monograph, *The Linguistic Basis of Articulation Rate Variation in Czech*, is perhaps the most comprehensive treatment focusing on articulation rate, or at least of the causes of variation in rate, in the literature. While her data are from Czech and she is careful to keep her observations and findings placed in terms of articulation rate variation in the Czech language, her consideration goes well beyond a specific language and has much to offer a general understanding of articulation rate. Dankovičová focuses on what she describes as developing a "theory of articulation rate" (2001: 5) centered on the question of whether articulation has a "domain" – "a unit within which articulation rate is organized" (2001: 23) – and, if so, what that domain might be. Her interests, as indicated by her title, are on the linguistic factors that influence articulation rate, such as word position, word length, and word class (i.e. content vs function), and she downplays social or speaker-based variation. While including her individual speakers as a factor in her analyses, she overarchingly attempts to limit between-speaker variability. She notes, "each speaker has his own characteristic overall articulation rate, which is, under comparable circumstances, relatively stable" (Dankovičová 2001: 112). Further, she argues:

> These results cannot provide very strong support to what seems to be a general belief, that people differ in how fast they speak. Of course, there are speakers whom we notice as particularly slow or fast, but these are likely to be at the edge of the spectrum ...; for most speakers, their articulation rates do not differ to a significant extent, at least not at the level of phonological word. (Dankovičová 2001: 132)

However, her study only examined seven speakers, and she selected speakers who were most similar to one another demographically and in terms of her perception of their speech rates. Her second experiment, for example, examined four speakers selected as the most stable from among ten speakers she recorded for the task. This seems to me more as evidence that we can successfully identify speakers who have comparable rates than as evidence that most speakers do not differ in terms of their general articulation rates.

Dankovičová (2001) provides a brief survey of a range of literature on the role of "independent variables" on rate variation. She briefly discusses the role of factors like age and gender, and then reviews studies which mostly focus on the role of task type on articulation rate. For instance, sports commentaries have been found to be spoken at a faster rate in Hungarian than poetry recitals (Fónagy and Magdics 1960) and at a faster rate than spontaneous speech in Czech (Bartošek 1974). Considering studies that have compared read speech to spontaneous speech, Dankovičová reports:

> The findings are contradictory. While Hewlett and Rendall (1998) reported spontaneous speech (conversation) being faster than reading, the opposite was observed, for instance, by Lehtonen (1979) and Butcher (1981). Butcher reported the value of 6.13 syll/s in reading and 5.26 syll/s in spontaneous speech but no significant differences between two types of spontaneous speech (reminiscence and retelling a story). A faster articulation rate in reading than in spontaneous speech was also found by Strangert (1993) for Swedish. (2001: 12–13)

In a more recent comparison, Jacewicz and colleagues (Jacewicz, Fox, and Wei 2010) found that rates in read speech (on a per-speaker basis) were correlated with, but less than, rates in conversational speech. Although I do not directly compare read speech to conversational speech in this work, the unclear relationship between the two seems important for any large-scale consideration of speech timing. I have more to say about this in Chapter 4 where I analyze pause and speech rate in recordings of reading passages.

Dankovičová's own empirical look at articulation rate variation involved two related experiments where she investigated rate variability

within different units of speech (intonation phrases versus phonetic utterances versus syntactic clauses) and what factors influence that variability. Her findings suggest that

> Articulation rate in elicited spontaneous Czech speech does not vary in an arbitrary way but has a domain within which it varies systematically. This domain is the intonation phrase and the pattern of variation is a slowing down throughout the phrase. The analysis also showed that the slowing down is nonlinear; it is rather gradual across non-final words, with the last word within the phrase being significantly slower than any other proceeding words within the same phrase. (Dankovičová 2001: 65)

While Dankovičová argues for the intonation phrase being the primary domain of articulation rate variability, she also notes that this domain correlates highly with phonetic utterances (what she calls INTERPAUSE STRETCHES) and, though to a lesser extent, clause boundaries. She further notes that the phrase-final lengthening effect is more pronounced at pause boundaries than at intonation phrase boundaries within larger interpause units (2001: 124). In addition to finding a robust slowing of rate over the course of intonation phrases, Dankovičová also found a strong effect of word size (in syllables) on articulation rate; longer words are produced faster than shorter words (e.g. 2001: 94–6).

One critical observation about Dankovičová's study, from the perspective of the current project, is – as is obvious from her title – that she was interested primarily in linguistic factors behind articulation rate variability. Her experiments, while quite detailed and extensive, examine three and four speakers respectively and also seek to limit the conversational naturalness of the spontaneous speech she examined. For instance, she explains,

> Regarding elicited spontaneous speech, I recorded between 7 and 14 minutes of speech per subject. My preference was to use parts of the speech which (i) were not interrupted by my questions ..., and (ii) which seemed most natural in the sense of matching closely the subject's speech outside the recording situation. These criteria compelled me to cut down the length of spontaneous speech samples to a maximum of about 4 minutes per speaker. (2001: 68)

She also chose subjects who were as similar as possible (young adult, middle class, standard Czech speakers) in order to limit social

variability. It is understandable why for her purposes – an interest solely in linguistic factors – she limited the speech in this way, yet these facts are clearly detrimental for drawing further conclusions from her work.

The work 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 existence and significance of speaker-level differences in speech rate. For example, Goldman-Eisler (cf. 1968) found that her subjects showed a great deal of individual differences in their overall speech rates, but 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) and argued that all of the 57 speakers he examined had quite similar rates. 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, and this seems 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 et al. 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 above the level of individual speakers have been examined to some extent, primarily in terms of regional differences. At a macroregional level, Robb et al. (2004), for example, compared speech rates between 40 speakers of New Zealand English and 40 speakers of American English and found that the New Zealanders had significantly faster articulation rates (and speaking rates) than the Americans, demonstrating that "not all varieties of English are spoken at the same rate" (Robb et al. 2004: 12). Regional differences in speech rate have also been found within American English by some researchers (e.g. Jacewicz, Fox, O'Neill, and Salmons 2009, Jacewicz et al. 2010), but not by others (e.g. Freiman 1979, Ray and Zahn 1990,[9] although Ray and Zahn noted surprise at their null result). The recent studies by Jacewicz and colleagues represent the most comprehensive looks at regional speech rate differences in the US to date. These studies are a part

of those researchers' attempt to characterize cross-generational change in dialect systems and are based on impressively large collections of read and conversational speech from southeastern Wisconsin, western North Carolina, and central Ohio (although most of the analyses have focused on the comparison between the Wisconsin and North Carolina speakers). While much of their research has focused on the dialect regions' vowel systems and not specifically on speech timing, the work specifically on speech rate has shown that the Wisconsin speakers speak significantly faster than the North Carolina speakers in both read and conversational tasks. This could, perhaps, be taken as some of the best available evidence in support of the popular stereotype of slower speech in the South. Yet, some caution is still necessary in drawing broad conclusions even from this dataset and thorough study. Western North Carolina is but one place in the American South and, in fact, is often considered to be different from the rest of North Carolina, let alone the "South" as a major dialect area. In fact, most linguistic research in western North Carolina (e.g. Mallinson and Wolfram 2002) discusses its variety as Appalachian English rather than Southern English. But this is somewhat digressive – the primary point here is that no studies (to my knowledge) have closely examined rate differences across communities or subregions within a single, larger regional variety. In Chapter 5, we will look closely at four different parts of North Carolina to ask how variable rates are within regions and not just between regions.

Investigations of sex-based variation in speech rate have also yielded conflicting results. Many studies have pointed to males speaking faster than females, but it is often weak or mitigated evidence. Yuan et al. (2006) found men speaking faster than women, but also noted that the difference between males and females, albeit statistically significant, was very minor. Jacewicz et al. (2009, 2010) found that males had significantly faster speaking rates than females, but that in read speech the differences were not significant. Deese (1984), on the other hand, found that the women in his data spoken significantly faster than the men. Ray and Zahn (1990) do not find significant differences by gender. Clopper and Smiljanic (2011), examining differences between Midland American English and Southern American English, also found no significant differences for gender (or dialect) on speech rate.

Age appears to be the social factor which has been studied the most in research on articulation rate, likely as a result of a wide interest in tracking first language acquisition (and fluency) in young children and, to a lesser extent, first language declination in aging populations. Here there seems to be fairly robust findings across studies indicating

a nonlinear change in rate over the course of speakers' lives. Children have the slowest rates, which increase over adolescence and peak in middle adulthood. Rates then appear to decrease as individuals move into older adulthood and old age. Yuan et al. (2006), Quené (2008), and Jacewicz et al. (2010) all provide evidence of this. Dankovičová (2001: 10–11) reviews other projects which support this summative view.

In sum, studies of speech rate have found significant differences at the individual level and between macroregional varieties. In terms of finer-level, sociolinguistically relevant differences, however, findings have been contradictory, with some researchers finding significant differences at the regional and gender levels (e.g. Jacewicz et al. 2010) and others finding no significant differences (e.g. Clopper and Smiljanic 2011).

One reason for the contradictory findings – beyond Goldman-Eisler's (e.g. 1968) suggestion that speech rate might be 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 – in his 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. Jacewicz et al., in their mixed-effect modeling look at their data (2010), included phrase length (in seconds of duration) and found it to be an important significant factor. In their case, however, the factors in the model including phrase length were quite similar in their effects to the model without a phrase length predictor. In other words, for their data, the phrase length effect does not appear to otherwise change the outcome of the model. In fact, their findings also run counter to Quené's (2008) with respect to the direction of the phrase length effect. They found that shorter length phrases have faster speech rates while Quené found longer phrases to have faster speech rates. It is hard to know offhand whether this difference is a result of differences between Dutch and American English or differences stemming from methodological decisions. To risk getting ahead of myself, in my analysis we will see very strong effects for phrase length, with rates increasing rapidly over the shortest utterances and then (somewhat) plateauing for long utterances.

Jacewicz et al. only included utterances with five or more syllables in their data and excised fillers from the speech. Quené also excised some of the shortest utterances – explaining "most of the short phrases

(of one or two orthographic syllables) consisted of hesitation sounds, filled pauses, backchannel sounds, etc. [and these] were excluded from the dataset" (2008: 1105). Jacewicz et al. (2010: 845–6) consider possible reasons for their contrary findings to Quené's (2008) and they mention the possibility of this relating to their decision not to include utterances shorter than five syllables. These decisions clearly influence the outcome and the comparability of both studies' findings. Fillers are often longer in duration, and hence slower than nonfillers, so removing these will likely have a nonlinear but increasing effect on the resulting speech rates. As we will see in Chapter 5, short utterances make up a huge proportion of talk in natural speech – 38 percent of the utterances in my dataset contain one to four syllables. The decision whether or not these are included in the analysis will surely have a large impact on the findings. Excluding these data and trimming out fillers seem to me unhelpful and artificial maneuvers. After all, in normal interactions we listen to speech with all of its *um*s and *uh*s.

Jacewicz et al. (2010) also model the influences on phrase length (in terms of duration in seconds) and find that Wisconsin speakers produce significantly shorter phrases than North Carolina speakers and that older speakers produce shorter phrases than younger speakers. They explain, "the significantly faster speaking rate [i.e. articulation rate] of Wisconsin speakers seems to be related to shorter phrases in their productions. By the same token, longer phrases produced by North Carolina speakers affect their speaking rate [i.e. articulation rate], which is significantly slower" (2010: 846). Dankovičová (2001) also found that utterance length had a major impact on articulation rate. In Kendall and Thomas (2010), Erik Thomas and I investigated the effect of phrase-final lengthening on articulation rate and I will return to our data and findings below (in §6.4).

A second reason for conflicting results in the previous literature may relate more simply to the varied measures (see §4.4) 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 differences) hide for some studies what might otherwise be found to be significant variation.[10] 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 remains at issue is that we might disagree with the categorization of a 1.5 σ/sec range as "a small range of variation." When considering speech features like speech rate and pause, we must revisit the discussions from earlier in this chapter and ask to what degree differences in these features

are perceptible to listeners.[11] As a reminder from §2.2, Quené (2007) found that hearers perceive rate of speech changes greater than about 5 percent (i.e. that the just noticeable difference, or JND, is ~5 percent). This indicates that hearers may perceive differences in speech rate on the order of ±0.25 σ/sec (based on an average speech rate of somewhere around 5 σ/sec). In other words, differences in speech rates between 4.4 and 5.9 σ/sec would be quite noticeable and should probably not be considered "a small range of variation" at all.

2.5 Motivating further study

It is of general interest to the recent expansion of sociophonetic pursuits, as well as to general sociolinguistic concerns, whether variability in pause realization and speech rate has significant/discernible social correlates. As should be clear from the paucity of direct sociolinguistic studies in the reviews of the above sections, very little work has systematically assessed this question (especially from the rich conceptual toolbox developed in sociolinguistics). However, it is plausible for a number of reasons that pause and speech rate have social correlates. There is vast evidence from morphosyntactic, phonological, and segmental phonetic sociolinguistic research that children learn the fine probabilistic patterns of their community of peers. Why would temporal patterns be any different? Even if pause realizations are so tied to cognitive and task-related factors that these reduce the available space for social differentiation, it still seems the case that there may be room for social patterns to emerge. As, for instance, Grosjean (1980a; Grosjean and Deschamps 1975) demonstrated in his comparison of pause patterns in French and English, language varieties may be able to distribute their pause time in systematic but different ways. If this is so, it will help us understand the extent to which language is socially influenced. If these features prove to be idiosyncratic or, more simply, chaotic and unpatterned, it will usefully indicate the fact that some features truly are not socially patterned. (In fact, as a hint of what is to come in Chapter 8, a finding congruent with the notion that pause durations are patterned in ways that result primarily from cognitive factors, like task difficulty, allows us to develop interesting sociolinguistic hypotheses about the use of pause as a potential predictor for other sociolinguistically relevant variables.)

Goldman-Eisler's experiments found, for pause, that "there were individual differences and characteristic ranges of [pause time]; individuals were consistent in their tendency to hesitate or utter speech fluently. We must therefore assume something like a characteristic disposition

to pausing" (1968: 68). For speech rate, she also wrote that "the rate of articulation is a personality constant of remarkable invariance" (Goldman-Eisler 1968: 25). Work over the past 40 years has indicated that although these earliest studies were quite valuable there is much more to pause and speech rate patterns than what they showed. Are these features truly personality traits? Are speech rates and characteristic ranges of pause time idiosyncratic? Or, do we find patterns of larger social differentiation when we investigate these features from the perspective of sociolinguistics. The popular belief in or even character of the "slow-talking Southerner," "the long-pausing Native American," and so forth, would indicate that rate differences exist somewhere in the social world. And, of course, the recent studies like those by Jacewicz et al. (2009, 2010) and Quené (2008), which found regional differences in these temporal patterns, indicate that there is much room for sociolinguistically meaningful patterns to exist for these features.

Further, Chafe's work, introduced earlier (e.g. 1980a, b), on timing and in particular on the role of pauses in the flow of consciousness in discourse, raises several further areas worth pursuing by sociolinguists. It is more than reminiscent of the attention to the speech model introduced by Labov (1966[2006], 1972) for dealing with intra-speaker variability, or SOCIOLINGUISTIC STYLE. And ultimately I will return to this towards the end of the book. But before turning to my substantive pursuits, there is more to say about the corpora, methods, and tools which background this project, and these are the topics of the next chapter.

3
New Tools and Speech Databases

3.1 Introduction

Before moving on to the actual empirical pursuits of this monograph, it is worth backing up a bit to the origins of this project, to its foundations in work on data management in sociolinguistics and in consideration of the nature of sociolinguistic data. The research discussed here was initially inspired by my ongoing work on methodologies for databasing and archiving speech recordings (Kendall, 2008a, 2009, 2010a, 2011, forthcoming a, b, Kendall and Bradlow 2011). Over the past half-dozen years, I have been involved in the development of two projects in particular, SLAAP and OSCAAR (described, and acronyms expanded, below), which center on the creation of web-based digital archives built around time-aligned annotation frameworks. Ultimately, it was the development of these time-aligned frameworks and an exploration of theories of transcription (Kendall 2005, 2006–2007), which led to my interest in speech timing phenomena. In a sense, I stumbled into questions about pause timing (Kendall 2006) as I explored various ways users might interact with the time-aligned transcription model implemented in SLAAP.

In my PhD dissertation (Kendall 2009) I took up many of the questions of this book – in an introductory fashion – as a supplement to the general description of the approach to sociolinguistic data and data management implemented in SLAAP. The work there examined about 100 speakers from the SLAAP archive to demonstrate the ways that "recycled" sociolinguistic data could shed light on new questions, questions which were not part of the original research projects that collected the speech recordings in the first place. As I stated in Chapter 1, pauses and speech rates are ubiquitous features of talk, and a large and growing

archive of spontaneous speech recordings, which share a fine-grained, time-aligned transcription model, seemed a great place to investigate the variability of speech timing phenomena.

While my goal here is not to stray too far into methodological issues in the management and preservation of sociolinguistic recordings (readers are referred to Kendall 2008a, 2011, forthcoming a, and b for more specific considerations of data and corpora in sociolinguistics), I review here the technical and methodological considerations that form the backdrop for the actual empirical analyses of Part II. First, in §3.2, I describe the Sociolinguistic Archive and Analysis Project (SLAAP), the home of the data and tools used for most of the analyses of this book. In §3.3, I discuss the time-aligned transcription system that forms the basis for the pause and speech rate measurements used as the data of this book. In §3.4, I very briefly introduce OSCAAR, a related archiving project which houses a smaller set of data examined in Chapter 4. Then, in §3.5, I return to SLAAP to describe more specifically the tools developed for the analysis of speech timing phenomena.

3.2 The Sociolinguistic Archive and Analysis Project (SLAAP)

The Sociolinguistic Archive and Analysis Project (SLAAP) centers on a web-based archive and analytic toolset for sociolinguistic data collections, but simultaneously encompasses a broader effort to explore new approaches to storing, managing, and interacting with natural speech data. SLAAP began in 2005 as a digitization and preservation collaboration between the North Carolina Language and Life Project (NCLLP), a research initiative at North Carolina State University,[1] and the North Carolina State University Libraries and was first envisioned as a resource specific to the NCLLP's materials. Over time, SLAAP has grown to become a more broadly used speech data management system and recording archive. SLAAP increasingly seeks to provide a central repository for sociolinguistic recordings from outside the NCLLP and is adding large collections of non-NCLLP materials. (While human subjects' considerations and agreements prevent fully open access to the archive, some materials can be shared with others for research purposes and researchers can request access to the collections in the archive following information on the main website – http://ncslaap.lib.ncsu.edu/.)

To a certain degree SLAAP looks like some of the other corpus development projects discussed in the recent literature (such as the ONZE corpus discussed by Gordon, Maclagan, and Hay 2007 and the LANCHART database discussed by Gregersen 2009). However, SLAAP seeks to fill a gap in terms of sociolinguistic practice more than it seeks to create a particular corpus (Kendall 2008a, forthcoming a, b). 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 with no envisioned end-product. It is a SPEECH DATA MANAGEMENT SYSTEM (SDMS), which is designed to house and organize an expanding collection of audio recordings. The archive is actively growing as part of our ongoing digitization and transcription effort. As of February 2012, the SLAAP digital archive contains over 2600 interviews and over 2100 hours of audio.[2] Over 50 hours have associated time-aligned transcripts, making a transcript collection of over 500,000 words.

The recordings housed in SLAAP share a metadata format and an underlying structure (in terms of how the structural elements of the data – such as speaker records, recording metadata, project-level information, etc. – are stored and linked) as well as transcription and annotation protocols, but they come from research projects spanning several decades, from audio cassette-based field recordings to high-quality digital recordings conducted in university settings. Most of the recordings in the archive are sociolinguistic interviews (cf. Labov 1972, Milroy and Gordon 2003), but other recordings, like those of public events or radio interviews, are also included when deemed relevant for sociolinguistic research and/or they have been collected as part of a sociolinguistic project.

The specific goals behind SLAAP are multiple. At a practical level, as mentioned, the project 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 through a variety of web-based features (cf. Kendall 2007a). 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 (cf. Kendall 2008a). As I hope the studies in this book demonstrate, 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 (which is described in the next section). SLAAP allows large-scale corpus-like sociophonetic analyses of timing patterns through highly accurate, instrumental techniques. With the tools developed in SLAAP, it is possible, as we do in Chapter 5, to extract

for analysis tens of thousands of speech rate measurements from the archive somewhat automatically.

By digitizing the entire NCLLP collection and incorporating the recordings into a centralized repository, we have in a sense put into dialogue numerous collections of sociolinguistic data. The descriptive metadata – i.e. the information stored about each interview, speaker, and research project – along with 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 and consistent 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. And, as the research here demonstrates, it can also create opportunities to evaluate new research questions. The analyses of Chapter 5 and the following chapters 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.

We now turn to discuss SLAAP's transcript model in more detail. In the following chapters, we treat the transcribed speech data as the corpus-based data for analysis. As such, the design of the data is a crucial part of the analysis and has ramifications on the types of questions that can be asked, and the possible answers obtained (Kendall 2008a, forthcoming b).

3.3 SLAAP's transcript model

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. Through SLAAP's software, the same transcript can be viewed in a VERTICAL FORMAT (as in (1) in Figure 3.1; Edwards 2001) or a COLUMN-BASED FORMAT (as in (2) in Figure 3.1; Ochs 1979, Edwards 2001), or even in what is referred to in SLAAP as a PARAGRAPH FORMAT (as in (3) in Figure 3.1).

Alternatively, the same transcript can be transformed in various ways, such as into purely visual formats. The view shown in (4) of Figure 3.1 (and in Figure 4.5 in the next chapter), called a GRAPHICALIZATION (Kendall 2007a), displays speakers' utterances within the complete interaction in

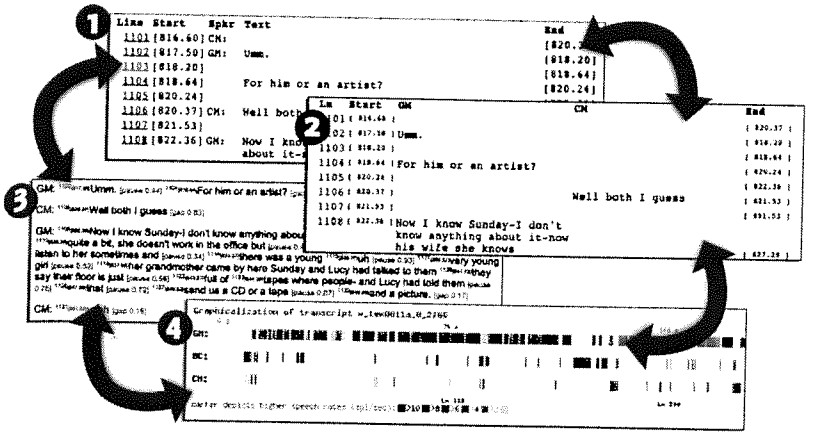


Figure 3.1 Four presentations available in SLAAP of the same transcript data (from Kendall 2007a)

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,[3] 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 3.3).

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 the TextGrid features of Praat (Boersma and Weenink 2010) to obtain highly accurate start and end times for each utterance.[4] Each speaker is orthographically transcribed in his or her own TextGrid tier so that the temporal record accurately records the times of that specific speaker's contributions. The central unit of the transcript is the PHONETIC UTTERANCE – a stretch of speech bounded by pauses. Pauses are delimited separately from the speech, with a 60 ms threshold used as the minimum silence captured as a pause.

Figure 3.2 displays the Praat Editor window for the same transcript displayed in Figure 3.1 above. This represents the "source" transcript before it is added to SLAAP. The example shows three utterances for the interviewee GM (the full text for the third utterance is shown by Praat although the actual audio, wave form, and spectrogram run off-screen to the right). The second and third tiers house the transcriptions for

the two interviewers BC and CM, although in the 8-second window shown only CM speaks, with a single utterance. The interval boundaries accurately capture the start and end times of each utterance and in doing so accurately delimit the pauses. In the Praat window shown, the 442 millisecond pause between GM's utterance "Umm." and "For him or an artist?" is selected. In SLAAP and the analyses of the following chapters, this silence is considered a pause because it falls between two utterances by the same speaker. The next silent interval by GM is not deemed a pause because CM speaks during the span of time. One of the benefits of recording individual speakers' contributions on their own tiers in Praat is that it allows for the accurate delimitation of the speech and a full accounting for each speaker over the entire course of the transcript. Overlap between two (or more) speakers is accurately captured, as the individual tiers will also show overlapping intervals when two (or more) speakers talk at the same time. (No overlapping speech occurs in the example of Figure 3.2.) While we are not focusing on GAPS, pauses between speaker turns (see e.g. Mendoza-Denton 1995) in the treatment in this book, gap lengths can be computed from comparing the end and start times of adjacent speaker turns. (For instance, the gap between GM's "For him or an artist?" and CM's "Well both I guess" can readily be computed from the two boundaries as 130 milliseconds.) While Praat can be used to examine and analyze aspects of the transcripts, for our purposes the transcripts are simply developed using Praat. From there, they are imported into SLAAP, where a software component of the archive processes the TextGrids and converts them to

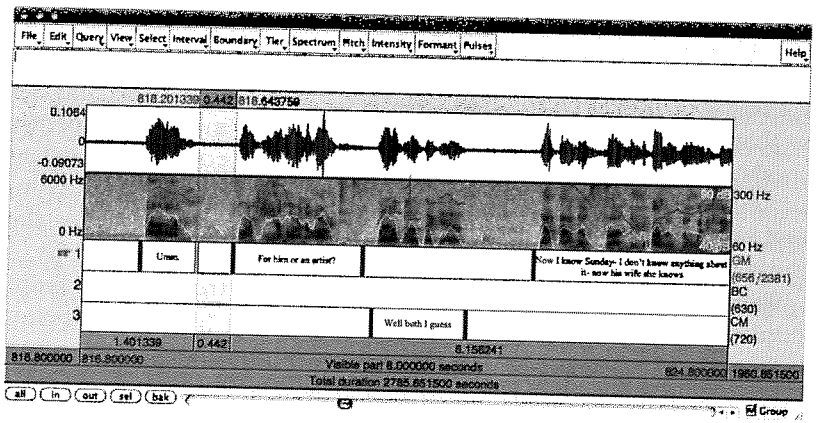


Figure 3.2 Praat TextGrid for the transcript shown in Figure 3.1

the data-based versions of the transcripts housed and accessed through the web-based software.

As this discussion illustrates, the fundamental components of SLAAP's DATA-BASED TRANSCRIPT MODEL are quite simple. In such a 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 2006–2007, 2007a, 2009). Through specially designed software, like SLAAP, this very simple data model is quite powerful. SLAAP creates links between the transcript data and the audio file from which the transcript is based, and phonetic software (such as Praat in the case of SLAAP) can be integrated into the transcript interface software to allow for real-time phonetic analysis from within the transcript. 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) is unnecessary and can be reconstructed from the audio itself.

At the same time, an approximation of standard orthography (cf. Chafe 1993: 34, Tagliamonte 2007: 211–15)[5] is sufficient for the transcript text because pronunciation features (e.g. vowel qualities, *r*-vocalization) 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 nonexperts. The use of standard orthography also allows for easier searching and for more straightforward concordancing and other corpus-based extraction measures (cf. McEnery and Wilson 2001, McEnery, Xiao, and Tono 2006). For the purposes of the studies of this book, the simple orthographic representation of the speech means that a fairly simple rule-based syllable-counting algorithm can dependably count syllables from the stored text (see §4.4.1).

As an illustration of what software can do with this simple, but data-based, transcript data, Figure 3.3 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 3.3 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 easy comparison between utterances.

With orthographic transcription data stored in a database and accurately time-aligned to the source audio, many transformations, manipulations,

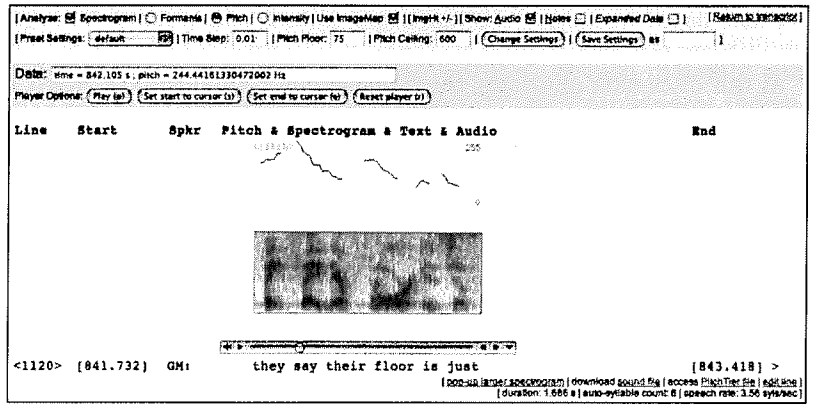


Figure 3.3 SLAAP screenshot showing a transcript line with phonetic data

and queries become available. I will return to discussing these possibilities and the specific tools available in SLAAP for speech timing research after briefly introducing OSCAAR, a project related to SLAAP.

3.4 The Online Speech/Corpora Archive and Analysis Resource

In addition to SLAAP, I have also been involved in the creation of an archive and speech data management system in the Linguistics Department at Northwestern University. This project, the Online Speech/Corpora Archive and Analysis Resource (OSCAAR; http://oscaar.ling.northwestern.edu/; Kendall 2010a, Kendall and Bradlow 2011), was begun as an expansion and generalization of SLAAP. OSCAAR seeks to extend SLAAP's approach to the storage, management, and preservation of speech recordings to a more diverse range of speech recordings, with a specific focus on the kinds of recordings generated in lab-based phonetics and speech science work.

While SLAAP focuses entirely on sociolinguistic data, and on rethinking how researchers can access and analyze those data, OSCAAR is designed as a more general repository with the bulk of its features centered on providing organizational aid for the large amounts (often thousands) of short recordings generated in lab-based research, such as the Wildcat Corpus of Native and Foreign-Accented English (Van Engen, Baese-Berk, Baker, Choi, Kim, and Bradlow 2010) and the LUCID corpus (London UCL Clear Speech in Interaction Project; Baker and Hazan 2010). OSCAAR also provides a set of tools to link source stimuli for production data, like sentence reading prompts, reading passages, images (e.g. the diapix scenes of the Wildcat Corpus and LUCID), and so forth to their derivative recordings. The highly specified transcription conventions developed for SLAAP are relaxed in OSCAAR, where the transcript-based features have been rewritten to provide generic, web-based access to Praat TextGrid files, rather than access and analysis tools to span across separate recording collections.

For our purposes, OSCAAR is similar enough to SLAAP to not warrant a longer discussion or screenshots of its own. I introduce it here primarily because it hosts the reading passage data used for the "first look" at speech rate and pause variability and the explication of my analytic methods in the next chapter.

3.5 Tools for the analysis of temporal speech features

As mentioned above, one of the major features of the SLAAP (and OSCAAR) 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.

SLAAP has a number of corpus-like, analysis features that automatically or semiautomatically extract features from the time-aligned transcript archive. The relevant tools will be introduced and discussed briefly here. In addition to the analysis tools, SLAAP provides a number of interfaces with the transcript archive. Some basic views of the transcripts were provided in the "collage" of Figure 3.1 and the close-up view of a single transcript line in Figure 3.3. To illustrate the organization of and interface to the transcript collection more generally, Figure 3.4 shows a screenshot of the SLAAP transcript summary list page. In this view, sets of transcripts (here, those associated with the Robeson County collection, from research in southern North Carolina; Wolfram et al. 2002) are available along with information about the speakers in the transcripts and their lengths. Links are available to various transcript-based features.

As an illustration of the sorts of information SLAAP can generate about each transcript, Figure 3.5 shows an excerpt from SLAAP's transcription

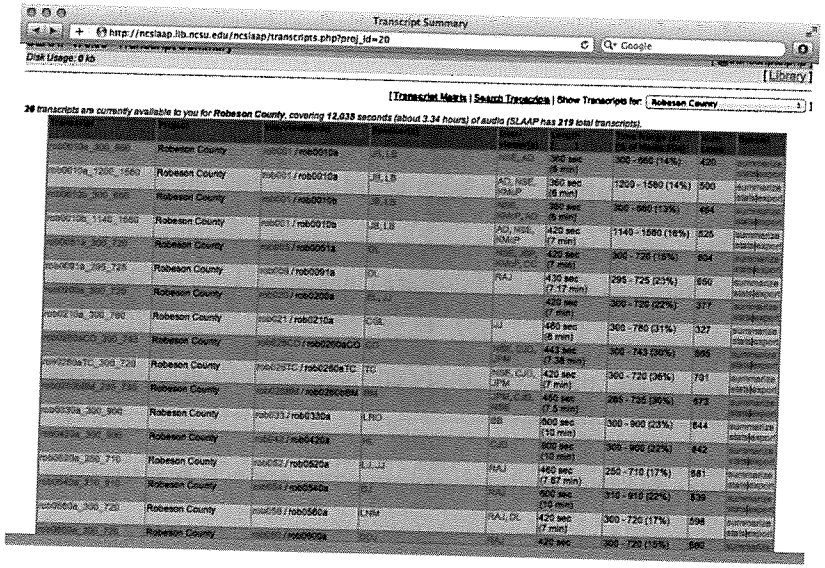


Figure 3.4 SLAAP screenshot of transcript summary list for Robeson County

Transcript Summary Statistics

Transcript has 2 speakers: *PE, ERT*

Transcript total temporal length: **911.00** seconds (15.18 minutes)

Transcript total line length: **1,208** lines (including blank lines, e.g., pauses)
Total non-blank lines: **603**

Speaker	Talk Lines[1]	Turn Lines[1]	Words	Words of Tran	Talk-Time (sec)	Talk-Time of Total Talk[2]	Turn-Time (sec)	Turn-Time of Entire Tran[3]
PE	453	786	2,936	86.76 %	588.45	86.79 %	754.91	82.87 %
ERT	150	187	448	13.24 %	89.56	13.21 %	110.33	12.11 %
Totals:	603	973	3,384	100 %	678.01	100 %	865.24	94.98 %

[1] **Talk Lines** only include transcript lines with orthographic text. **Turn Lines** are all transcript lines that occur within a speaker's turn. The crucial difference between **Talk Lines** and **Turn Lines** is whether or not blank lines, or pauses, are counted. Blank lines are determined to "belong" to the speaker by occurring between two lines of talk. **Talk-Time** and **Turn-Time** are sums of the timespans of these two measurements of line "ownership".

[2] **Talk-Time of Total Talk** is the percentage of total talk (not including pauses) by each speaker. The sum of all the speaker's **Talk-Time** should always account for 100% of the total talk in the transcript.

[3] **Turn-Time of Entire Tran** is the percentage of how much of the entire duration of the transcript's time each speakers' total **Turn-Time** accounts for. The sum of this measure will usually be less than 100% as not all lines (namely, inter-turn pauses) "belong" to specific speakers. A high amount of Speaker of overlap (more overlap than inter-turn pause) can result in a result over 100%.

Figure 3.5 Excerpt of SLAAP screenshot showing summary statistics for the transcript for media file ptx0120b

summary statistics page for the transcript of media file ptx0120b, an interview from a community in southern Texas by ERT with PE, as the interviewee is labeled. This view gives us summary information about the selected transcript, including information about the total contributions to the talk by each participant.[6]

Other examples of SLAAP's general purposes features – including its audio player and extraction features and non-transcript-related analysis and research features – are described elsewhere (cf. Kendall 2007a, 2008a, 2009). We continue to focus on the transcript-related features and turn now to the most important features for the analyses of the coming chapters, the corpus-like speech timing analysis tools. Figure 3.6 shows a screenshot from SLAAP of the speech rate analysis page. This feature, based on user-specified settings (such as the range of utterance durations considered and the maximum number of utterances to retrieve) extracts individual utterances from the specified transcript (excluding utterances containing unsure transcription, speaker overlap, or nonlinguistic noises) and calculates a syllable count and a syllables per second articulation rate measure for each matching line. The syllable counter is described in §4.4.1 and its code is provided, as ported to a function for the R language, on the book's website.

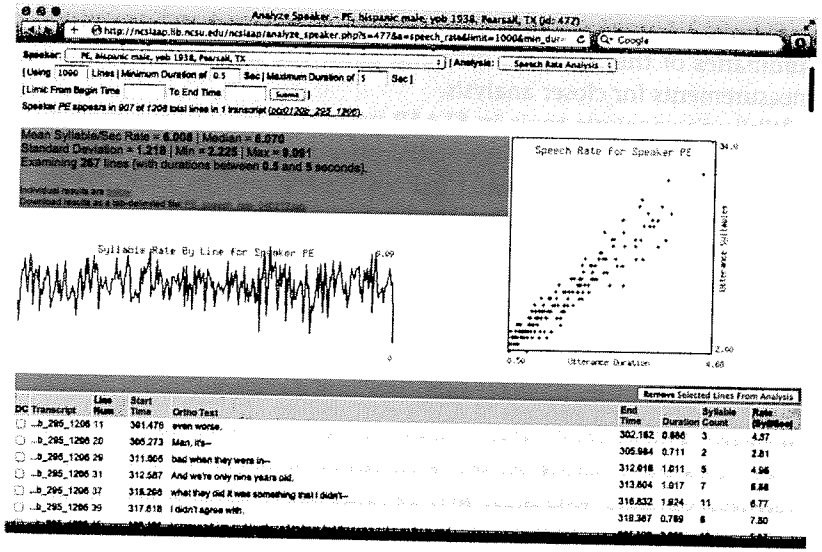


Figure 3.6 Screenshot of SLAAP's speech rate analysis tool

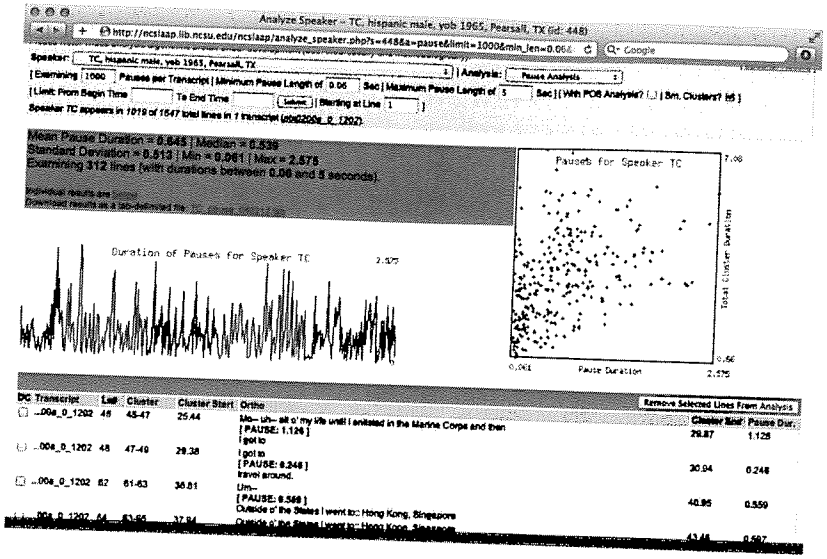


Figure 3.7 Screenshot of SLAAP's silent pause analysis tool

Figure 3.7 displays SLAAP's pause analysis feature. This tool, again based on criteria set by the user, finds and extracts all matching pauses that are bounded by uninterrupted talk by the same speaker. As is visible in Figures 3.6 and 3.7, both of the tools provide visual and quantitative summaries of the extracted data and allow the user to download the measurements for closer analysis.

While the analysis tools in SLAAP provide easy-to-use interfaces to extract speech timing data, the actual extraction of the data for a large-scale analysis is still somewhat tedious if generated by hand for each individual of interest, as this involves configuring and running the analysis independently for each speaker and then downloading each set of results and compiling them into a larger spreadsheet of data. For the large-scale analyses of Chapter 5, I have written scripts in the programming language R (R Development Core Team 2011), which communicate with the SLAAP server and extract the necessary data. These scripts 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 quantitative and statistical analysis. It is these files of compiled, extracted data from the many transcripts that form the measurements analyzed in Part II, the empirical studies of pause and speech rate.

Part II
Studies in Speech Rate and Pause Variation

4

Methods and a First Look at Speech Rate and Pause

4.1 Introduction

Here I present the first of several empirical investigations into speech rate and pause variation. In this first view, I examine a small experimental dataset drawing from read speech from three regions of the US. These data consist of recordings of a short reading passage read by 14 talkers from Memphis, Tennessee (the South), 14 talkers from Oswego, New York (Labov et al.'s 2006 Inland North), and 14 talkers from Reno, Nevada (the West). The talkers are all natives of their respective larger dialect regions, are European Americans, and are adults in the 18–30-year-old range. The data are drawn from research with Valerie Fridland (Kendall and Fridland 2012, Fridland and Kendall 2012).[1] Unlike the data from the remaining chapters, which all come from SLAAP's archives, the recordings examined here are stored in OSCAAR, the web-based speech resource archive housed at Northwestern University, which was introduced briefly in §3.4.

This analysis serves as a prelude to the main analyses of Chapter 5 and I use it largely as an opportunity to review the major methodological issues involved in these kinds of analyses. I also hope the discussion here, and throughout this book, highlights some of the possibilities that exist for the analysis of temporal phenomena in human speech and can lead readers to their own, new inquiries about these kinds of features. I begin in §4.2 by discussing the statistical methods used throughout this book, and then briefly describe the data used for this preliminary analysis in §4.3. I move on to a more general consideration of technical aspects of the analysis in §4.4, like the syllable-counting algorithm used here and throughout this study. This section also includes a discussion of the major methodological issues surrounding the study of speech rate and pause, including the

important question of what kinds of units of measurement are most useful for looking at variation in these temporal features (this revisits some of the discussion from Chapter 2 on terminology and measures of speech rate). In §4.5, I examine in some detail the speech rate and pause duration patterns that emerge from these reading passage data. Finally (and as a preview of the results of this preliminary analysis), in §4.6, I argue that read speech is somewhat problematic for assessing social variation in speech timing partly on the grounds that subjects in laboratory settings read in idiosyncratic ways, which likely affects aspects of timing and ultimately confounds an analysis looking for social, group-based differentiation. Further, read speech has temporal characteristics that do not seem equivalent to naturally occurring conversational speech. These facts, I propose, point towards the value of using collections of recorded conversational interview speech for the sociolinguistic study of temporal phenomena.

4.2 Modeling sociophonetic data

In the last chapter, I framed the projects of this book around my recent work on the management and annotation of speech data, pointing out the ways that the software and transcription system in SLAAP makes conducting corpus-like studies of sociophonetic questions available for a diverse set of existing data. It is also the case that the present project is made more possible by recent advances in the statistical analyses available for linguistic research. While I do not intend this book to act as a full-fledged primer in statistical analysis for sociophonetic/corpus phonetic research, I do take advantage of the book format to provide longer descriptions of the statistical analyses than are often given in most journal articles or research reports. In fact, as statistical analysis makes up a large component of the remaining chapters, I begin by providing some background into some recent advances in statistics that are available for linguistic research. I hope that these discussions are helpful for readers, both for understanding the specific steps I have undertaken and for informing other large-scale sociophonetic research. For readers interested in actual guides for the kinds of statistical methods I use in this book, and mixed-effect modeling in particular, I recommend Harald Baayen's (2008) *Analyzing Linguistic Data*, which forms the basis for much of my use of regression and mixed-effect modeling. For a more sociolinguistic and historical focus, Daniel Ezra Johnson's (2009) *Language and Linguistics Compass* article provides a nice review of statistical methods in sociolinguistics and a compelling argument for the use of the mixed-effect approaches I use here.[2]

Variationist sociolinguistics has long centered upon statistical analysis. One of the discipline's defining features has been the near ubiquity of the use of the VARBRUL program (Cedegren and Sankoff 1974) over the past 30 or so years of research. Varbrul was developed in order to provide a form of LOGISTIC REGRESSION designed specifically for the characteristics of sociolinguistic data. Varbrul (and logistic regression in general) tests the impact of statistical predictors, also called independent variables, on the realization of categorical dependent variables, the features of interest. Thus, Varbrul (and, again, logistic regression in general) can test questions like: What are the factors that significantly impact whether a word ending in *–ing* was realized as *–in'* or *–ing*, what is commonly called VARIABLE (ING). This analysis holds up even when tokens – individual data points – are unevenly distributed across the independent variables. Many independent variables can be tested together, and each can have their own arbitrary number of levels. Thus, an analysis of (ing) can test a range of factors at once (such as linguistic factors like the phonological environment preceding and/or following the *–ing*, the grammatical status of the form in question, and so forth, and social factors, like speaker sex, social class, age group, and so on).

For several decades, Varbrul and its descendents, like GoldVarb (Sankoff, Tagliamonte, and Smith 2005), were the cutting-edge statistical tools available for variable sociolinguistic data. And, until recently, very few alternatives to Varbrul had been available or used by sociolinguistic researchers. In addition to only working for categorical dependent variables (like *–ing* vs *–in*; a property of logistic regression more generally), Varbrul and its descendents have some limitations. Importantly, they can only test predictors that are categorical as well – so age, for instance, or the height of a vowel, must be binned into levels in order to be included in an analysis – and they cannot test for interactions between factors. Sociolinguistic variable data (cf. Wolfram 1993, Tagliamonte 2006) have tended to be categorical and Varbrul's focus on categorical variables was primarily a design feature. (See Tagliamonte 2006 for an excellent overview of GoldVarb and more general variationist data analysis procedures.) However, recent sociolinguistic – and in particular sociophonetic – inquiries have increasingly been interested in continuous data.

For continuous data, like acoustic measurements (and, for example, speech rates and pause durations) and much psycholinguistic data, ANOVA (ANalysis Of VAriance) has been the most widely used statistical method in linguistics. Put simply, ANOVA tests whether the means of different distributions of data, organized by categorical factors, are different or not. As such, ANOVA are similar to *t*-tests (the simplest test of whether

two groups of data have the same or different means) but designed to test larger sets of groups than just two, and more complicated organizations of data. Statistically speaking, many types of ANOVA are a kind of LINEAR REGRESSION. Linear regression – the method used for much of the analyses in this book – is similar to logistic regression, as I have described it above, except that linear regression tests continuous dependent variables rather than categorical ones. Linear regression, as a general statistical approach, can be used to test whether groups of data have significantly different means (just like ANOVA) or can be used to predict estimated values for a dependent variable, given a set of predictors. Further, linear regression tests the effects of continuous and categorical predictors, while ANOVA tests only categorical predictors. Logistic regression is in fact a special case form of linear regression, where instead of the statistical model estimating actual values for the dependent variable the model estimates a predicted likelihood of the categorical factor outcome, typically through a LOG-ODDS value – literally, the log of the odds of an outcome (cf. D. E. Johnson 2009). Historically, linear regression and ANOVA have not been used all that often by variationists; even when collecting continuous data, for example vowel height, variationists have traditionally tended to collapse continuous features into categorical variables in order to be able to use Varbrul, the preferred statistical method.

In recent years, advancements in general statistical analysis have become available for linguistic research, and the more general regression techniques I have just described have rapidly gained in popularity. While the techniques implemented in Varbrul were ahead of their time (especially in terms of their availability and accessibility to linguists), regression modeling has become a ubiquitous form of statistical analysis across disciplines and is implemented in ways useful for linguistic analysis in numerous software packages. Many of these packages have surpassed Varbrul's abilities to accurately model sociolinguistic data (cf. D. E. Johnson 2009). These general-purpose regression methods allow researchers to model all kinds of data, without having to coerce the data to fit the software. Independent variables that are both categorical and continuous can be included in the statistical models, as can interactions between factors (i.e. whether one factor's effect on the dependent variable depends on the state of a second factor). For instance, extremely powerful general logistic regression and linear regression libraries are available in the open-source programming and statistical environment R (R Development Core Team 2011).[3] The available R packages, like the Design (Harrell 2009) and rms libraries (Harrell 2011), the lme4 library (Bates and Maechler 2010), and the languageR library (Baayen 2008),

all provide helpful tools for doing statistical analyses and, further, for conducting MODEL CRITICISM, the determination of just how "good" one's statistical models are.

In particular, MIXED-EFFECT MODELING methods (in contrast to FIXED-EFFECT MODELING, as the traditional regression methods are termed) have been developed which allow regression procedures to more accurately model complex data with different kinds of influences, like sociolinguistic data: individual data tokens come from specific speakers and, typically, multiple tokens are gathered from each individual in a study. Tokens collected or measured from the same speaker are not independent of one another and thus break the formal, theoretical assumptions of many statistical techniques. Further, while we might be interested in, say, the importance of regional affiliation as a category impacting the production of a particular variable, different speakers who share the same regional identification will both simultaneously conform to their group norms and exhibit idiosyncratic traits. Also, different individuals will contribute different numbers of tokens to the analysis, so these idiosyncratic differences can have unbalanced influences on the overall dataset and the outcome of statistical models that do not account for the individual differences. Mixed effects allow the statistical models to adjust for these facts.

Mixed-effect models have so-called FIXED EFFECTS (like sex or social class or grammatical category and so on) for which the full space of possible values is known. Fixed effects can be modeled to determine how significant and how strong the effect is – the same as in traditional forms of regression. Some factors involve too many possible levels (like speakers or words), however, where one cannot readily enumerate (or sample) all of the different values. Put differently, these are factors for which the values in one's data can be thought of as a random sample of the possible values. These are the RANDOM EFFECTS, which the model can adjust to account for individual differences. Thus, by accounting differently for the random effects, mixed-effect models allow for the generation of simultaneously more accurate and more theoretically appropriate models. They do come at a cost, however, in that mixed-effect models involve more complicated mathematical formulae and do not yet have as established or straightforward techniques for model criticism. Standard regressions (i.e. fixed-effect models), as well as recent Varbrul programs, readily provide information about how well a given model accounts for the variability in the data, but less straightforward means are needed to do similar work for mixed-effect models.

Another aspect of mixed-effect modeling has to do with the assumed predictive capabilities and the replicability of the statistical models.

Models using random effects are built around the fact that the individual random effects (for our purposes, the individual speakers from whom the data come) influence the data. If one were to replicate a study, one could readily obtain new data with the same fixed effects, but, presumably, the random effects would not be replicated (for example, one would most likely not have the same speakers as subjects). Mixed-effect models then better account for this important difference in effect types, and, with the proper techniques, can be more accurately and more theoretically appropriately put to a range of applications (such as predicting the realization of a dependent variable in unseen data).

With advantages, and some disadvantages, over traditional statistical techniques, mixed-effect modeling is rapidly becoming a dominant method for the analysis of data like those in this book. Many good sources are now available that discuss using mixed-effect models for linguistic research, such as Baayen (2008), Jaeger (2008), K. Johnson (2008), and Quené and van den Bergh (2008). As I mentioned above, D. E. Johnson (2009) discusses mixed-effect modeling specifically in terms of *sociolinguistic* methodology and analysis practices and compares these techniques to Varbrul analysis. His paper also introduces Rbrul, a Varbrul-like program implemented in R that offers both the traditional features of the Varbrul programs and new features, like mixed effects and predictor interactions. Beyond linguistics, other volumes – such as Pinheiro and Bates (2000) and Luke (2004) – provide general introductions to mixed-effect methods, and a much larger literature (and online communities and help fora) are available for fixed-effect regression modeling. I discuss some aspects of conducting mixed-effects modeling at further length in §4.5.1 and §5.3.1 when these methods are first used in each of these two chapters.

4.3 The reading passage data

The 42 talkers examined here all read a short story of 266 words (cardinally 324 syllables). The reading passage was as follows:

> *Some mornings in the summertime, when the sky is fair and the lawn covered in dew, the good Duke Post and his wife Peg walk down to the brook by their house. There, beside the trees, is their favorite place to sit, talk and sip coffee. Her father, Don, and his dog, Bookie, often stop by to chat while their children, Betty and Kate, toss off their shoes and leap headfirst into the deep brook. It makes Peg feel like a kid again to watch them dive, shout and slosh around in the water and swing off the old black tire tied to the oak tree.*

> *One hot hazy, dull afternoon, she gave a call to their friends Pam and Ben Powder, inviting them over for supper. On the way, their truck got stuck in the mud and they showed up an hour late, for which they caught a good deal of teasing. But soon the crowd was having fun and the good hosts put out tunafish sandwiches, hot dogs, a big pot of bean soup and beer bread. When they were done eating, it was a sin that no one had saved room for Peg's tasty spice cake that was yet to come.*
>
> *After supper, Duke, Ben and his pal Bill went out on Duke's inflatable boat. Unfortunately, the sky got dark and started to pour rain. Bill lost his footing on the slick bank and fell in the water. After ten minutes he finally got into the boat. Once back on shore, the sudden weather shift sent everyone home, and the party was over.*

The recordings were time-aligned to the reading passage using the Penn Phonetics Lab Forced Aligner (P2FA; http://www.ling.upenn.edu/phonetics/p2fa/; Yuan and Liberman 2008) and were then post-processed using custom scripts in Praat (Boersma and Weenink 2010) so that the passages were aligned at the utterance level with all silences greater than 200 ms delimited in the text. Finally, the aligned prose was hand-corrected to match the actual productions of the talkers, so that each time-aligned file reflected an accurate orthographic transcription of the actual speech of the reading, and not simply the cardinal, read text. For illustration, Figure 4.1 displays a screenshot of a Praat Editor window showing one of the reading passages with its time-aligned text.

All silent pauses (again, longer than 200 ms) were isolated from the speech. Any silent stretches in the recordings before each subject

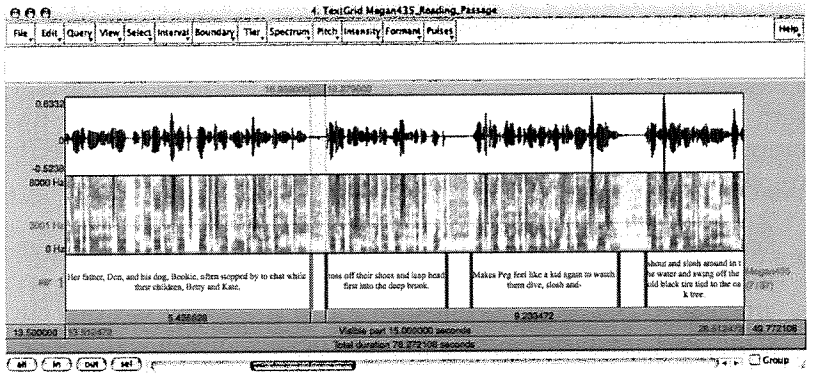


Figure 4.1 Praat Editor window showing one of the reading passages

began his or her first utterance and after his or her final utterance were removed from the data, but otherwise all silent intervals, regardless of their syntactic location (i.e. placement with respect to grammatical phrases or clauses) or prosodic position (i.e. placement with respect to prosodic factors like intonation) are included in the analysis. While the reading passage contained 324 syllables as written, it was produced with as few as 322 syllables and as many as 346 syllables (median 327.5), due to restarts and reading errors. As mentioned earlier, it should be remembered that the approach used to store, generate, and analyze these read data is slightly different than that used for the conversational, spontaneous speech examined in the rest of this book.

4.4 Measuring and defining rate of speech and pause

These reading passage data are partly examined here as an opportunity to review several methodological issues that arise when examining rate of speech and pause, whether in read or spontaneous speech. For instance, there are numerous ways to measure a speaker's rate of speech and to describe these rates. The decisions made while measuring and analyzing the data can have far-reaching implications on the outcome and its comparability to other studies. Before moving on to discuss the actual analysis, it is helpful to compare the range of available methods and to understand the distributions that we find in these sorts of data. Along these lines, I here outline and justify the decisions made for the analyses of this chapter and throughout much of the rest of the book.

4.4.1 Rate of speech

What I have been calling in general terms SPEECH RATE is often decomposed into two separate measures, SPEAKING RATE and ARTICULATION RATE, in the literature. In these more precise terms, speaking rate is used to refer to a measure that includes pauses, while articulation rate refers to a measure with pauses longer than a certain threshold omitted. While these two measures are obviously related, they are not always or necessarily directly correlated. In his *Principles of Phonetics*, John Laver explains,

> Several different relationships between articulation rate and speaking rate are possible, depending on the continuity of speech. A fast articulation rate could be combined with a fast overall speaking rate if the speech is fluent, without frequent or long inter-utterance silent pauses. A fast articulation rate could be combined with a slower overall speaking rate if the speech is interrupted, with frequent or long inter-utterance silences. A relatively slow articulation rate could be part of an overall fast speaking rate, if combined with unusual fluency. There seems, at least in the English-speaking world, to be no necessary tendency for articulation rate and speaking rate to share the same tempo category. Goldman-Eisler (1968: 24), ..., showed experimentally that while speaking rate is positively correlated with the proportional duration of silent pauses in the speech material, speaking rate and articulation rate have no significant correlation. (Laver 1994: 541)

Two additional comments are in order. The first is that articulation rate, as a measure of rate over uninterrupted speech, can be computed based on single, phonetic utterances, or as a measure based on the total talk time (excluding pauses) of a passage. Speaking rate, on the other hand, since it includes silent pauses, is always computed based on larger stretches of talk. Second, while according to Laver there is no *necessary* correlation between the articulation rate and speaking rate measures, it may still be the case that these two measures are correlated in much actual speech. This seems to me an empirical question to take up in the analysis and we will examine how correlated speaking rate and articulation rate are for these reading passage data. For now, we note that these two measures are calculated in different ways and may result in different outcomes. Note that measures of speaking rate, as they include pauses, will necessarily be "slower" (in the sense of showing less units of speech over units of time) than articulation rate measures for the same material.

In this chapter, since the data are segmented at pauses longer than or equal to 200 ms, articulation rate refers to a measure of rate that excludes any silence of 200 ms or more. Many studies which use articulation rate as their measure of speech rate do not actually provide information as to the threshold used for removing silences from the data. This clearly can have repercussions on the comparability of results across studies. Here, the inclusion of silences shorter than 200 ms in the speech data means that articulation rates are necessarily slower than they would be if shorter pauses/silences were excluded. (In Chapters 5 through 7, articulation rates are measured from data with pauses as short as 60 ms excluded.)

Throughout the studies in this and the following chapters, I often use the term "speech rate" to refer to articulation rate, as this will be the major measure of interest after this chapter, or to refer to the concept of rate in general terms. When considering speaking rate, the pause-inclusive

measure, I will specifically term this speaking rate. I will occasionally use RATE OF SPEECH as a more intentionally neutral term, to avoid indication of a particular quantitative metric.

Beyond the issue of whether to include silence or not, rate of speech measures are often discussed in terms of some unit of speech over some unit of time. WORDS PER MINUTE (wpm) is a common metric, as is SYLLABLES PER MINUTE (spm or σ/m), but it is not entirely uncommon to see measures such as phones per second discussed. 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). (I will return shortly to the question of how syllables are measured.) The syllables per second measure provides a more precise unit of measure than words per minute, and helps to indicate the higher degree of accuracy available with modern techniques than was available earlier. A number of other scholars (e.g. Clopper and Smiljanic 2007, Hewlett and Rendall 1998, Jacewicz et al. 2009, 2010, and often Miller et al. 1984) also use this unit. Robb et al. (2004) provide a nice survey of the calculations used in previous projects.

Since rate of speech is concerned with the relationship between utterance temporal durations (what I will label as the variable UTTDUR) and their length in syllables (NUMSYLS), a characterization of a speaker's rate can also be formulated as the regression coefficient from a simple linear regression testing UTTDUR as a function of NUMSYLS. Put differently, if we envision a scatter plot of UTTDUR vs NUMSYLS for a given speech sample (or speaker or community, etc.), the slope of the best-fit line through the data points (SYLSLOPE) can provide a measure of the general rate of speech for that sample (or speaker or community, etc.). This formula is: UTTDUR = α + β * NUMSYLS, where α = intercept (the value of UTTDUR when NUMSYLS is 0) and β = SYLSLOPE. It is demonstrated in Figure 4.2 for Al4552 and Ab503, one of the Northern speakers and one of the Southern speakers, respectively, who are examined below. As can be seen in the figure, the slope of the best-fit line nicely captures the general trend for each speaker. The *r* value in each of the plots' legends indicates the tightness of the correlation between the two dimensions; values so close to 1 indicate very high correlations.

To the best of my knowledge this conception and measure of rate has not been pursued in the previous research on rate of speech. It seems to me to offer some advantages over the typical proportional measures of rate, although it does have disadvantages as well, in that it is less intuitive and harder to immediately translate a slope value for a speaker to a sense of his or her articulation rate in a more traditional measure. It

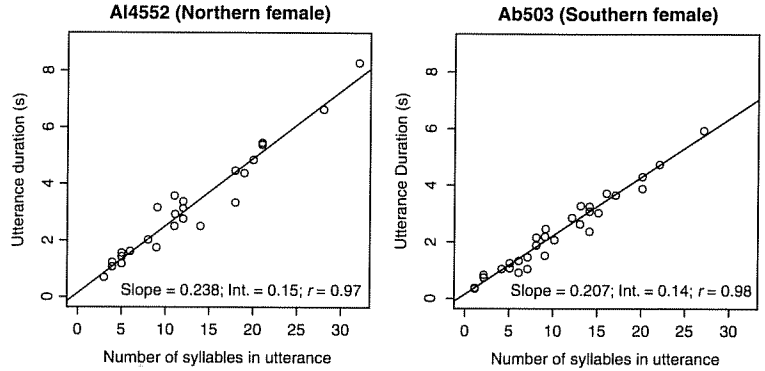


Figure 4.2 Considering rate of speech as a slope line

also has the potential disadvantage that the slope measure is a global measure over numerous utterances and cannot be computed for single utterances. I mention SYLSLOPE here primarily to indicate the range of possible ways that we can envision and characterize rate of speech. In order to keep my results as comparable as possible to the existing literature I do not pursue SYLSLOPE as a dependent variable in my analyses; I stick with the more conventional syllables per second measure.[4]

Some scholars (e.g. Tsao and Weismer 1997) discuss speech rates in terms of average syllable durations (ASD), using units like MILLISECONDS PER SYLLABLE (ms/σ). However these measures are presented, they are interchangeable, or at least should be, since the difference in measures is simply algebraic. Jacewicz et al. (2010: 845–6) tested this, by comparing their results using σ/sec against results using ms/σ and showed that the statistical results were the same. But, again, this should be the case, since the difference is simply a matter of which measure, syllable count or temporal duration, is the denominator, and which is the numerator. Slight differences may occur due to, say, how values are rounded, but large differences that arise due to choices of units should only occur through error. ASD strikes me as a useful measure *if* one is interested in examining the role of individual segments' or syllables' durations on overall rate of speech (or in using rate as a way to normalize other phenomena, like voice-onset-time measurements), but we will not be pursuing an interest in the durations of phonological units smaller than utterances (other than intonational phrases and their final feet in §6.4) as that would take us too far afield of the main interests of this work. Other work (in particular Dankovičova 2001) has investigated rate

variation within utterances and interested readers are referred to that work for more thorough discussions of utterance-internal variability.

In terms of actually measuring syllables in a given utterance, it is worth noting that many research reports on speech rate do not discuss their methods for counting syllables in any detail. In fact, while normal listeners (i.e. nonlinguists) are often quite good at identifying and counting syllables, defining a syllable is a notoriously difficult task (cf. Redford 1999, Ladefoged 2006). There are three main ways one can count syllables in a stretch of speech. One is acoustic, based on peaks in the acoustic signal (e.g. De Jong and Wempe 2009). One is auditory, based on an impressionistic counting of syllables while listening to the audio. And one is based on an orthographic representation of the speech. Each technique has advantages for different kinds of research questions. The first method – based on acoustic syllable detection – is probably the least commonly used method in the literature, while counting, automatically or manually, from an orthographic transcript, is likely the most common. Again, since many research reports do not go into the details of their syllable counting procedures, it is hard to know just how common each technique is.

The syllable counts used in the studies here are derived from the orthographic transcripts and generated from an automated algorithm. The algorithm used for all of the analyses in this book is from the syllable counter in the SLAAP software. It is a simple, rule-based counter that operates by first checking the input against a short list of lexical exceptions (e.g. "family" is coded as a lexical exception with two syllables and is not submitted to the main parser). Almost all words pass through the exception checking and then are parsed by the algorithm, which counts clusters of orthographic vowels in character strings (e.g. "deeper" is counted as two syllables, for "ee" and "e") and examines each word for pattern matches based on English spelling conventions (e.g. "hungrier" is counted as three syllables, one for "u", one for "ie", and one additional syllable because "rier" matches a pattern in the "add a syllable" list). Early versions of the algorithm achieved accuracy rates of around 80–85 percent but I have since improved the counter to the point where it yields close to 100 percent accuracy in tests, provided it is fed standard English orthography. (Improvements at this point are mainly made by adding a new pattern to the list of patterns or by adding the word to the look-up table of lexical exceptions as they are encountered.) While the algorithm will still make errors, any errors will be systematic across similar text strings in the counter's input and will occur throughout the data, thus limiting the impact of the errors, since

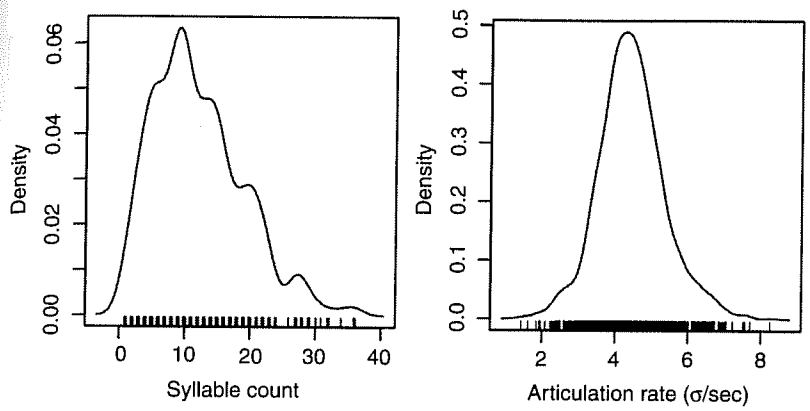


Figure 4.3 Syllable count and articulation rate measurement distributions

the data remain comparable across all speakers. Further, throughout all tests, even early tests on a less accurate counter, less than 2 percent of the incorrect counts are off by more than one syllable per utterance. The syllable counter has been ported to an R function and is available for download from the book's website.

Figure 4.3 displays the distribution of the syllable counts per utterance (on left) and the articulation rate measurements (on right) generated for the reading passage data. The mean number of syllables per utterance for all the reading passage data is 12.04 syllables, with a median value of 10.0 syllables.[5] The mean articulation rate is 4.44 syllables per second with a median of 4.38.

4.4.2 Pause durations

Pause durations are more straightforward to measure than speech rates in that we need to identify each silent pause and then simply measure its duration. Here, a pause is defined as a silence of greater than or equal to 200 ms (and as less than or equal to 5000 ms, although this high end is purely a formal definition here as the longest pause in the reading passage is 2684 ms, well below this maximum). Beginning in the next chapter, we will shrink the required minimum duration of a pause to 60 ms and, in §6.3, consider whether different duration cutoffs impact the outcome of a pause study. For the reading passage data in this chapter, any silent interval (greater than or equal to 200 ms) that occurs after the subject's first utterance of the reading until the last utterance is counted as a pause. When we examine conversational speech beginning

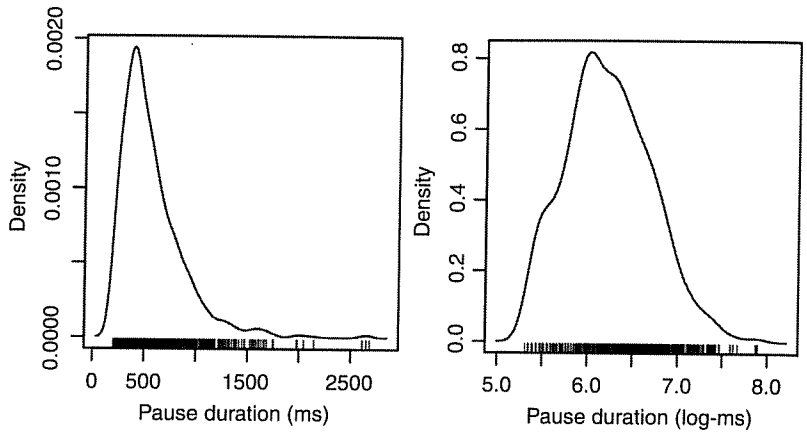


Figure 4.4 Pause duration measurement distributions (ms and log-ms)

in the next chapter, we restrict the pauses of interest to those that occur between spoken utterances within a speaker turn. Since there are no relevant interlocutors during a reading task, all silences – beyond disruptions or distractions in the experimental recording (none of which occurred for these data) – are deemed pauses attributed to the speaker.

Pause durations, since they limit 0 ms (or in this case 200 ms), distribute in a log-normal fashion. Figure 4.4 displays the distributions of the reading passage pause duration data, in ms (on left) and transformed into log-ms (on right). We will often examine pause durations transformed into log-ms, especially when modeling the pause duration data. Appendix II provides a simple conversion table between ms and log-ms.

4.5 Reading passage data and analysis

Table 4.1 displays summary data for the 42 talkers from the three regional locations by region and sex. The data for each individual speaker are available as Table 4a (as a Microsoft Excel formatted file and a tab-delimited text file) on the website.

The tables provide several measures for each speaker group (Table 4.1) and individual speakers (online Table 4a), including the median articulation rates and the overall articulation rates. The former is computed as the median value of the articulation rates of each of the speakers' utterances. The latter is a single computation, the total number of syllables spoken by that person divided by their total talk duration (available in

Table 4.1 Reading passage summary data

Speaker group	Median articulation rate	SylSlope	Overall articulation rate	Overall speaking rate	Total syllables	Median syllables per utt.	Pause *N*	Median pause dur. (ms)	Total pause dur. (s)	Total talk dur. (s)	Total reading dur. (s)
North females	4.38	0.1963	4.42	3.74	327.71	12.57	24.86	517.43	13.90	74.71	88.61
North males	4.36	0.1936	4.40	3.58	329.00	10.00	31.57	487.29	17.46	75.26	92.87
All North	4.37	0.1949	4.41	3.66	328.36	11.29	28.21	502.36	15.68	74.98	90.74
South females	4.31	0.1986	4.33	3.58	327.33	11.17	25.33	533.17	15.29	76.18	92.03
South males	4.23	0.1993	4.28	3.51	329.00	10.69	28.63	540.00	16.94	77.56	94.77
All South	4.27	0.1990	4.30	3.54	328.29	10.89	27.21	537.07	16.23	76.97	93.60
West females	4.41	0.1961	4.47	3.79	327.00	12.56	25.25	490.19	13.39	74.04	87.42
West males	4.98	0.1732	4.97	4.18	327.67	15.42	20.67	542.33	12.62	66.16	78.78
All West	4.65	0.1863	4.68	3.96	327.29	13.79	23.29	512.54	13.06	70.66	83.72
All females	4.37	0.1969	4.41	3.71	327.05	12.17	25.14	511.55	14.10	74.87	89.13
All males	4.48	0.1899	4.51	3.72	328.24	11.81	27.33	523.10	15.88	73.54	89.57
All talkers	4.43	0.1934	4.46	3.71	327.64	11.99	26.24	517.32	14.99	74.20	89.35

the second-to-last column). These two measures of articulation rate do not yield identical figures but are quite similar. On average, the utterance-derived medians are 0.03 σ/sec slower than the overall articulation rates, though some speakers have slightly faster median rates than overall rates so the (minor) differences do go in both directions. A *t*-test indicates that these two measurements are not significantly different ($p = 0.72$). The analyses, here and in the following chapters, will only examine articulation rates as measured per individual utterance (and also as collapsed as median values by speaker). Overall speaking rate is calculated by dividing the total number of syllables spoken by the total reading duration (the last column in the tables). The tables also include the SylSlope measure of speech rate introduced earlier in case readers wish to compare this measure with the others (SylSlope and median articulation rate correlate highly, and in an inverse fashion, although there are some differences and it is my hope that future work will investigate this measure more thoroughly; $r = -0.87$, $p < 0.000001$); I do not consider SylSlope further in the analysis. The remaining columns of the table should be self-explanatory.

As a further introduction to the reading passage data, Figure 4.5 displays GRAPHICALIZATIONS for the beginning portion of six of the talkers' readings. This technique was introduced briefly in Chapter 3's discussion of SLAAP (see also Kendall 2007a; in this case, the graphicalizations were generated by the OSCAAR web-based software, but the underlying software is almost identical). It presents a graphical timeline of a stretch of talk which displays talk as shaded rectangles whose width depicts the temporal length of each utterance and whose shading depicts the rate of speech for that utterance. The orthographic text for each utterance is displayed below it in the figure. Pauses are indicated as blank sections separating shaded rectangles. These graphicalizations are meant here as a way to provide a quick, qualitative overview for some of the data. Through them, we see examples of the variability across talkers in terms of how the reading passage is "chunked" in production and where pause time is distributed. Through the shading we see a coarse measure of where the subjects are reading faster and where they are reading slower.

We turn now to analyses of the speech rate data, and then pause data, from the reading passages.

4.5.1 Rate of speech in the reading passage data and its statistical analysis

Throughout these studies I will examine the dependent variables (i.e. speech rate and pause duration) at both a per-speaker (one central

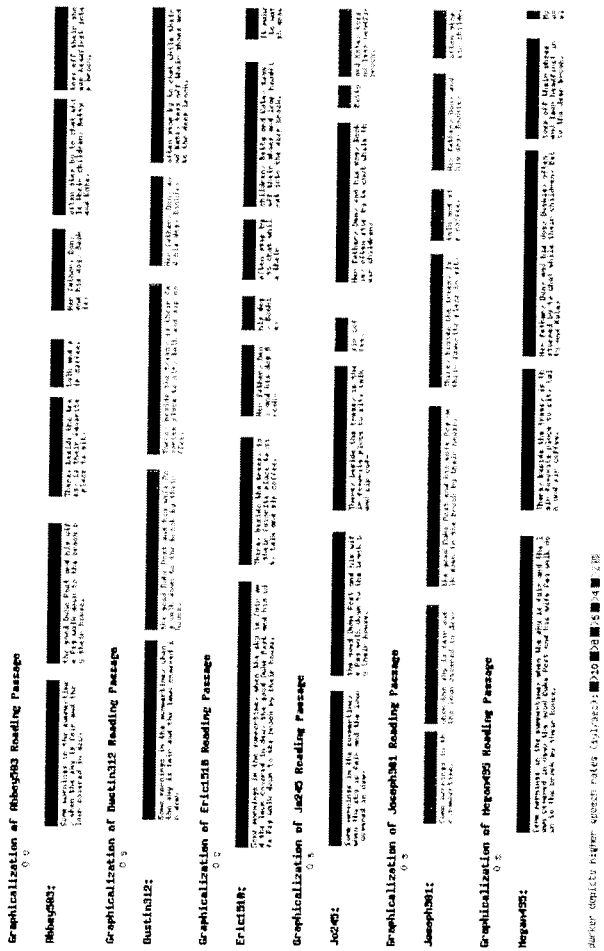


Figure 4.5 Graphicalizations of the beginning of six reading passages. The first two talkers are Southerners, the second two are Westerners, the last two are Northerners (talk extends off page to the right)

tendency per speaker) and per-measurement level (as many individual measurements as available for each speaker). Figure 4.6 displays boxplots[6] of the articulation rate measures collapsed over utterances by talker (right-hand panel), with 14 data points per region, each a talker median, and by individual utterance (left-hand panel). The two views of the data paint quite similar pictures, with the per-talker data showing the same general tendencies, although – as we would expect from a summary based on median values per talker – with less variation among the data. The Western talkers in both cases have the highest articulation rates, while the Southern talkers have the lowest rates, although the articulation rates for the Southerners and Northerners overlap to a large degree.

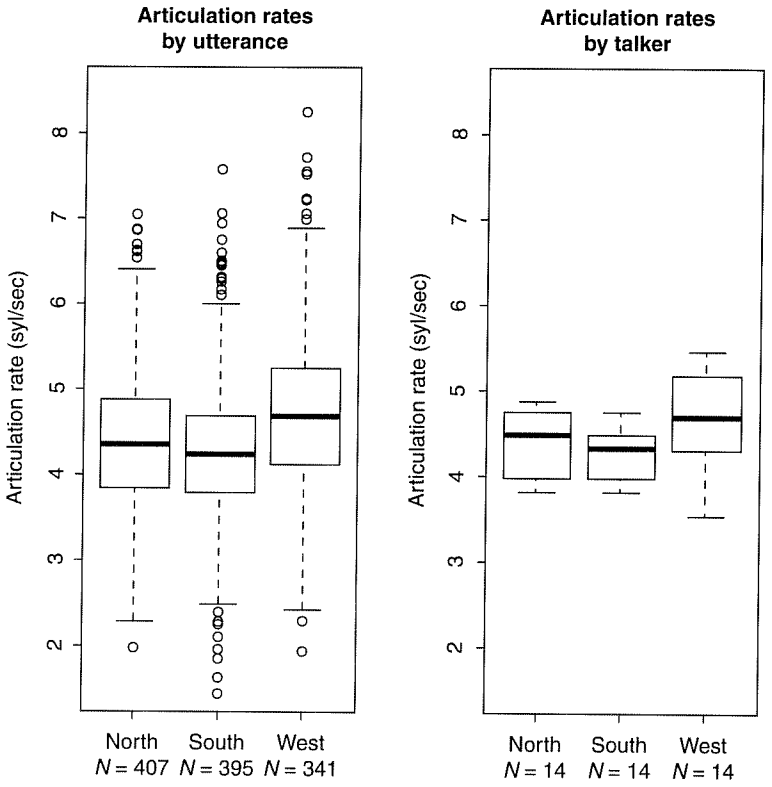


Figure 4.6 Articulation rates for reading passage data by utterance (left panel) and by talker (right panel)

As explained above in §4.4.1, we can also consider rate of speech in terms of the pause-inclusive measure of speaking rate. Figure 4.7, which shows the same articulation rate data by talker (left) along with the speaking rate measure for each talker (right), indicates that the articulation rate and speaking rate measures provide a similar comparison across the regions. In fact, the data points for both measures are highly correlated (Pearson's $r = 0.89$, $p < 0.000001$). Speaking rates are slower, of course, than articulation rates, because they represent the same number of syllables computed over (i.e. divided by) a longer duration.

From both Figures 4.6 and 4.7, it appears that Westerners have the fastest rates while Southerners have the slowest. To some extent, this seems like a reasonable finding, and is in line with the sorts of expectations gained from other studies of regional variation in American English speech rate, such as Jacewicz et al. (2009, 2010) and the folk

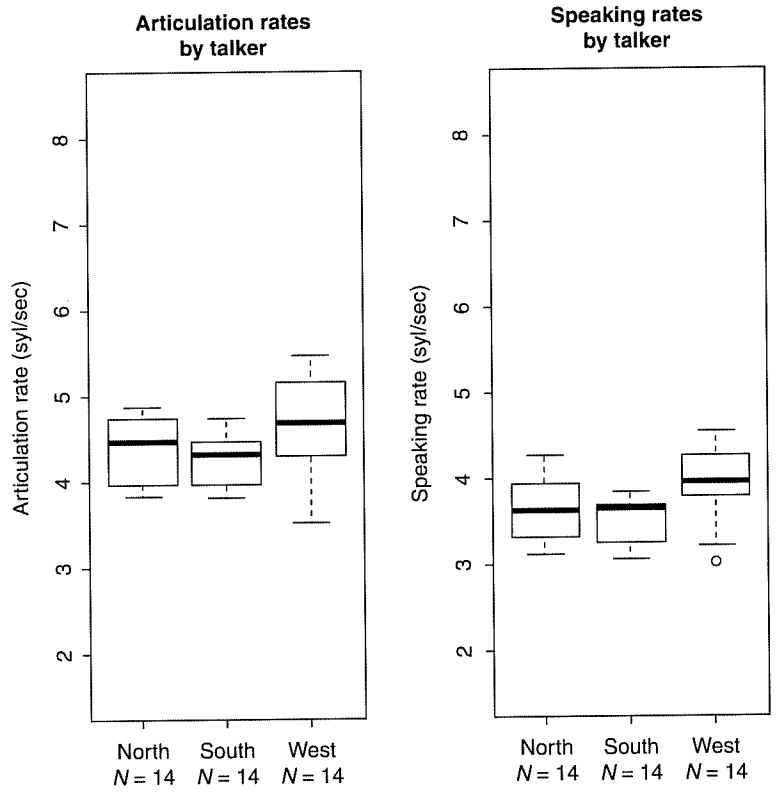


Figure 4.7 Articulation rates by talker (left) and speaking rates by talker (right)

notions of speech timing discussed in Chapter 2. At the same time, most of the academic and popular discourse on regional differences in speech timing indicates that Southerners are especially slow talkers. The picture here, however, is primarily that Westerners are fast talkers. The Southerners, while having slightly slower rates than Northerners (as indicated by the boxplots), are really not all that much slower.[7]

An ANOVA on the per-talker articulation rate data indicates that region is just significant ($F(2, 39) = 3.26$; $p = 0.049$),[8] but a Tukey post-hoc test shows that it is only the West–South comparison that is significant (and only at $p = 0.046$). An ANOVA for speaking rate also finds region to be significant ($F(2, 39) = 5.02$; $p = 0.012$) and again a Tukey test indicates that this significance is driven entirely by the West–South comparison ($p = 0.010$). The fact that the ANOVA for speaking rate has

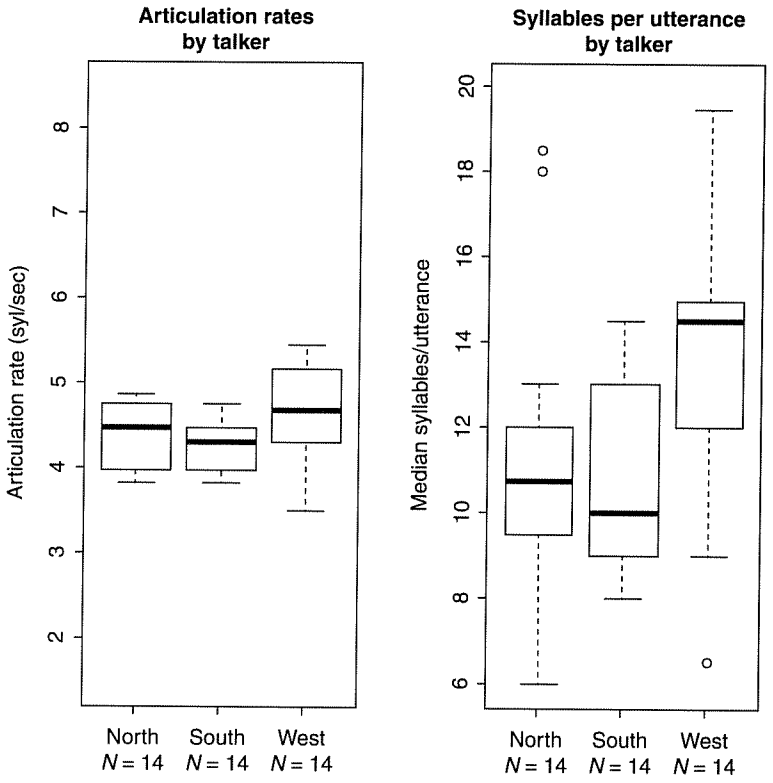


Figure 4.8 Articulation rates by talker (left) and median syllables per utterance by talker (right)

a lower p value and higher F value than the ANOVA for articulation rate possibly indicates that pause differences help to regionally differentiate the speakers, since speaking rate as a measure includes pauses. However, an interpretation of the patterns here based on these data alone is overly simplistic, and the picture gets more complicated when we look closely at the data in terms of the other factors present.

Figure 4.8 plots the same articulation rate data by talker (left) next to the syllables per utterance medians for the talkers (right), and here we see that Westerners, in fact, also have the longest utterances in addition to the fastest rates. Southerners and Northerners have overlapping median syllables per utterance measures but we also note that Southerners altogether have a lower central tendency. Region is found to be a significant factor in the difference in syllables per utterance in a simple one-way ANOVA ($F(2, 39) = 3.52$; $p = 0.039$). A post-hoc Tukey test here indicates again that the significant difference is also driven by the West–South comparison, which is the only significant comparison (at $p = 0.049$). In fact, a simple linear regression model testing the relationship between syllables per utterance and articulation rate, in turn, finds that median syllables per utterance significantly predicts articulation rate (at $p = 0.014$, with an estimated 0.5 σ/sec per syllable).[9] Thus, it is not clear yet whether the rate differences are simply a result of regionally different utterance length tendencies or whether the articulation rate differences are regionally different in their own right. We will return to this momentarily when we turn to a multivariate analysis.

First, we also ask of these reading passage data whether talkers' rates change in any systematic way over the course of the reading. Figure 4.9 illustrates that we in fact do find a systematic pattern over time; most speakers speed up over the course the reading passage.[10] This figure plots the articulation rate for each utterance (y-axis) according to that utterance's start time in the recording (x-axis) for each individual subject. In the captions for each plot panel, the "N," "S," and "W" denote Northern, Southern, and Western subjects respectively. Looking closely at the individual plots, and at the lowess lines in particular, which trace the overall tendency for each speaker, we see two main patterns. Some speakers have a slight general increase in their rates over the course of the reading and some speakers show a large "acceleration" in the last third or fourth of the reading. This is a striking pattern, and one that I will argue at the end of the chapter is *not* a characteristic of normal, conversational talk and is problematic for using data like these to formulate understandings of speech rate (in normal talk). Yet, this pattern makes sense in the context of read, laboratory speech. Subjects speed up over the course of the

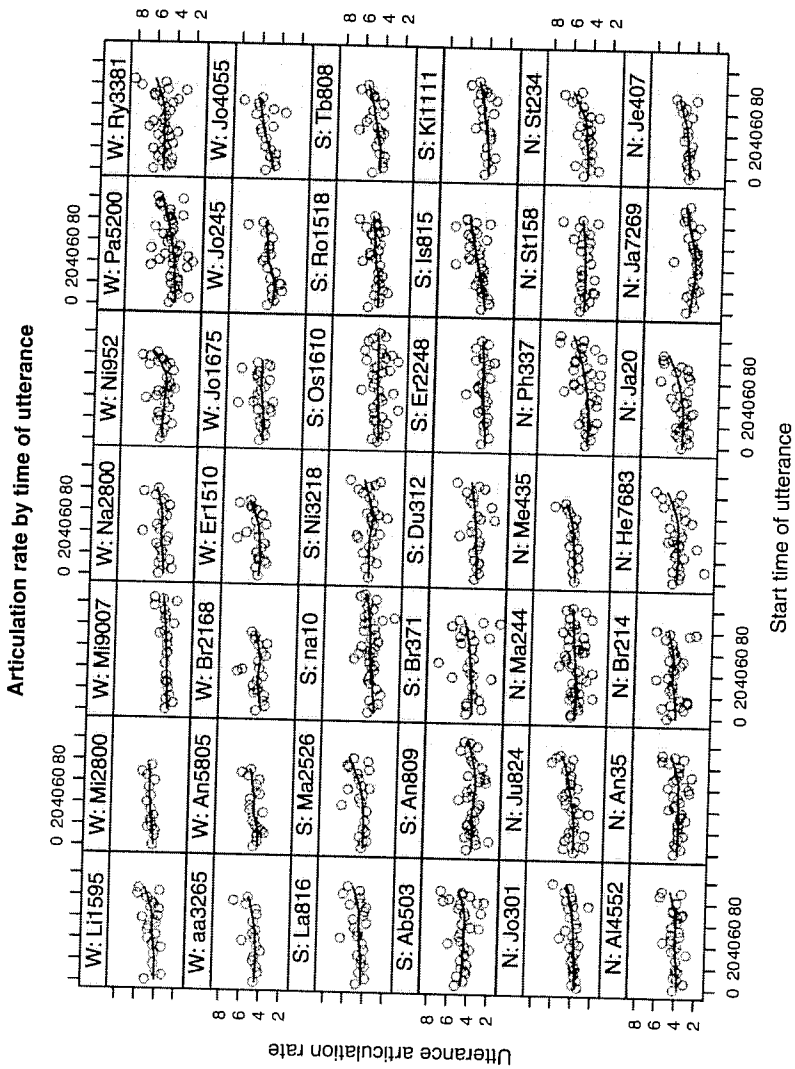


Figure 4.9 Articulation rates by utterance time for each talker

reading, and – although this must remain speculative without a more well-suited methodology such as eye-tracking – seem to speed up even more when they can tell they are approaching the end of the task.

We now turn to regression modeling, examining the full range of potential factors on the utterance-level measurements of articulation rate. As I explained in §4.2, a number of tools are available for mixed-effect modeling in R and I make heavy use of these tools in this book. For the most part, the approach to modeling I take follows Baayen's (2008) description of mixed-effect modeling for linguistic analysis. Unlike regression models used in traditional hypothesis testing (common in psycholinguistics), the modeling used here follows general sociolinguistic practice and is primarily exploratory. It tests the effects of all combinations of factors on an outcome in order to determine what set of factors best account for the variance in the data.

In developing the statistical model, I begin the analysis with the most basic possible model. In this case (and throughout this book) that is an intercept-only model with a single random effect (also called a random intercept) for speaker. This basic model only accounts for the individual speakers' different baseline rates and is used as a starting point to determine the first, most influential factor on the data. Modeling then proceeds by testing different fixed effects as potential factors that improve the model. The modeling for these articulation rate data examines the potential factors of REGION, speaker SEX, the number of syllables in the utterance (NUMSYLS), and the START time of the utterance. Each of these potential factors is tested and the one that significantly improves the fit of the model the most is then added to the model. This is repeated for the remaining factors until no further factors are found to improve the model fit.

During this process, some continuous factors, like the number of syllables (or age, although age is not examined in this chapter), where we might expect the factor to have a nonlinear influence on the dependent variable, are tested to see whether adding nonlinear components further improves the model. These nonlinear components can involve fitting quadratic polynomials to the model or, as is used commonly in this book to model the influence of the number of syllables in an utterance on speech rate, RESTRICTED CUBIC SPLINES, which allow the "lines" of fitted continuous predictors to bend in more flexible ways (cf. Harrell 2001, Baayen 2008: 174–81). Also during this process, other potential random effects are tested. In the speech rate and pause analyses I conduct, the only additional random effect that is tested is a random slope for the number of syllables in an utterance for the speech rate models. This

random slope can capture the fact that individual subjects can have somewhat idiosyncratic syllable length to articulation rate relations and the random effect helps the model control for these different individual tendencies. This random slope sometimes improves the models, and is included in the final model when that is the case, and sometimes does not. In other mixed-effect models of other kinds of linguistic data, the specific word that contains the phenomenon of interest can be a relevant random effect and is often tested.

Comparisons between the possible models are done by way of a likelihood ratio test, a statistical test that compares two similar models and indicates whether the more complicated model (the one with more parameters) is significantly better than the simpler model (cf. Baayen 2008: 253–4). Each added parameter, whether a factor or a nonlinear component of a factor already added to the model, adds complexity to the model, and the likelihood ratio test helps to indicate whether the benefit to the model is worth the added complexity. That is, whether the increase in the model's ability to account for the variance is enough to warrant the added "cost" of the new additional parameter. The process continues, slowly building up a larger model until no more factors are determined to significantly contribute to the improvement of the model. At this point, all the main effects are (typically) added and then various interactions are tested for in a similar fashion, one by one assessing which new interaction (if any) best improves the model. There is no absolutely straightforward rule for what order to add parameters and how to build models, and sometimes the addition of an interaction causes an earlier main effect to drop out of significance and this can then involve backing up several steps to determine whether the interaction or the previous main effect is in fact a more judicious inclusion in the model.

Once a full model is developed, the task then turns to model criticism and validation to ensure that the model is in fact fitting the data in a meaningful way – for instance that the model's residuals are normally distributed – and that it is not overfitting the data (cf. Baayen 2008: 188–95). For many of the mixed-effect models presented in this book, I also follow Baayen's (2008: 256–7) advice and trim outliers from the data, those data points with standardized residuals greater than 2.5 standard deviations from 0, after determining the most likely best model. I then refit the model on the trimmed data and report the final model after trimming. Trimming the data in this way removes the data points that are most unlike the other data points and this seems like a beneficial maneuver in developing an accurate sense of the trends in

the majority of the data. This trimming typically removes just under 2 percent of the data, although in the case of the reading passage articulation rate data, it removes 33 data points, or 2.9 percent, of the 1143 total measurements. The *p* values reported for factors in the mixed-effect regressions for articulation rate and pause duration are generated from the posterior distributions of a 10,000-iteration Markov chain Monte Carlo sampling method using the languageR library's `pvals.fnc()` function (Baayen 2008: 248).

Returning to the data at hand, the fixed effects for the best model for the articulation rate data, after trimming and refitting, are presented in Table 4.2. This model finds significant main effects for NumSyls, fitted with a six-knot restricted cubic spline, the Start time of the utterance, and Region. Speaker Sex is not included in the model as it was not found to significantly improve the model fit. In other words, Sex was found not to be a significant factor. As mentioned above, subject is included as a random intercept. Here the model fit is not improved by adding a random slope for NumSyls, so this is not included in the final model. NumSyls and Start time are both continuous predictors and thus normally have one coefficient, the estimated effect of a one-unit increase in that factor on the articulation rate of a given utterance. The six knots of NumSyls, however, result in the model having five coefficients for the factor, each of the additional four estimates representing an additional nonlinearity in the factor (as seen in the leftmost plot of Figure 4.10, shown below); these nonlinear components are noted in Table 4.2, and in later statistical results tables, using ' marks (thus NumSyls' reflects the first nonlinear coefficient, NumSyls'' indicates the second, and so on).

Table 4.2 Best mixed-effect model for (trimmed) reading passage articulation rate data

Factor	Estimate	Std. err.	*p*
(Intercept)	2.458	0.168	–
NumSyls	0.382	0.029	0.0001
NumSyls'	−6.402	0.676	0.0001
NumSyls''	16.182	1.922	0.0001
NumSyls'''	−13.393	1.980	0.0001
NumSyls''''	4.805	1.154	0.0001
Start	0.008	0.001	0.0001
Region = North (not West)	−0.288	0.161	0.0362
Region = South (not West)	−0.435	0.161	0.0024

$R^2 = 0.461$.

Table 4.2 (and the other linear regression model results in this book) displays the model's estimates for each factor's influence on the dependent variable, here articulation rate expressed in syllables per second. Normally, these tables can be interpreted by "plugging in" values of interest for the factors and adding the relevant estimate values to the intercept in order to extrapolate an estimated articulation rate. Throughout the analyses of this book, we will be examining these kinds of statistical models and some further comments on how to interpret these results are probably helpful. I will discuss how the model results for continuous factors (here, START and NUMSYLS) are best interpreted first, and then move on to discuss how categorical factors (here, REGION) are interpreted. The model results are also presented graphically in Figure 4.10.[11]

For continuous factors, the estimate is the change in the dependent variable per one-unit change in the predictor. So, for instance, a 1 sec increase in the utterance START time predicts a 0.008 σ/sec increase in articulation rate. For each model, I also show the model predictions in a graphical format, which can be an easier way to interpret the relative effects (and effect sizes) of the factors. In fact, interpreting the effect of factors modeled with splines, such as NUMSYLS, from the model results table is quite difficult and readers are urged to ignore those estimates in the table and use the graphical representation, here in Figure 4.10, to understand the nonlinear effect of utterance length.[12]

REGION, with three categorical levels, "North," "West," and "South," is modeled using so-called DUMMY, or TREATMENT, CODING. This is a typical way that categorical independent predictors are modeled and involves using simple binary comparisons rather than multiple-leveled data to build predictions. For these factors, one level is selected as the baseline level and then the others are compared, in pairwise fashion, to that baseline. R automatically dummy codes categorical variables and, by default, sets the baseline factor to the factor that is alphabetically first. The baseline can also be manually set to a specific factor, if this is desired. For these reading passage data, since we have seen evidence that the West is different from the North and South but that the North and South are not all that different from one another, I have set the baseline factor to be the West. Thus, the model results in Table 4.2, above, display the effect on articulation rate of the North in comparison to the West and of the South in comparison to the West. So if we are interested in estimating or predicting the rate of a Westerner, no value is added or subtracted from the other factor estimates. If the speaker is a Northerner, we subtract 0.288 σ/sec from the estimated rate and if the speaker of interest is a Southerner we subtract 0.435 σ/sec. In order to determine whether the

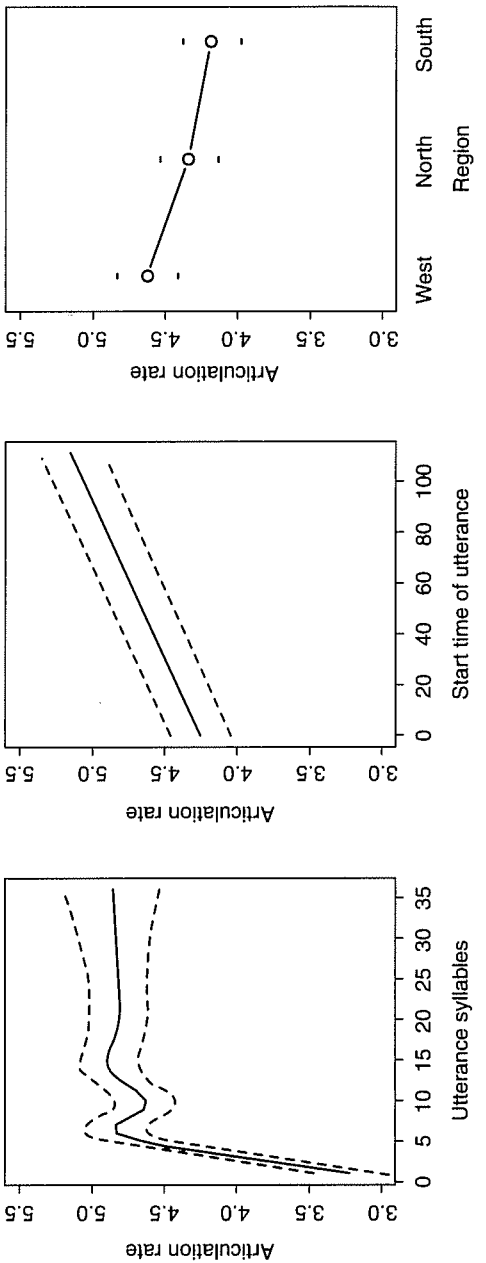


Figure 4.10 Effects in the mixed-effect model for reading passage articulation rates

third possible comparison (the North vs South) is significant, the REGION factor must be releveled with either the North or South set as the baseline and the model run again. In general throughout this book, I do not generate these additional versions of the models, unless the particular additional comparison is important. In this case, the additional model finds that the North vs South difference is not significant, with $p = 0.28$. (Further, REGION is not significant at all if the West is removed from the data and only the North vs South is tested, with $p = 0.21$.)

While the difference between the West and the two other regions is significant, we also note, especially from the graphical display in Figure 4.10, that the effect of REGION is actually not all that great. In fact, comparing the effect of REGION with the two other significant factors (even just by visually comparing the range of values on the *y*-axes of the figure), we note that REGION has the smallest effect size (and, in Table 4.2, has the highest – i.e. the least significant – *p* values). Nonetheless, the fact that REGION does arise as significant even when the model contains the number of syllables in each utterance shows that there are real regional differences in the articulation rate data. The differences by region are not simply a result of different utterance length tendencies.

The number of syllables in an utterance, NUMSYLS, has a massive effect, with a large range for the shortest utterances but then a mostly flat effect on utterances longer than about 15 syllables. The dip that occurs at around 10 syllables is hard to explain. It is perhaps a result of the prosody of the specific utterances falling in that length range. That is, since the data come from reading passage speech and all of the speakers read identical passages, many utterances cover about the same amount of the passage and it seems possible that some of these utterances in the 10-syllable range are biased towards slower rates by their syntactic structure or discourse (i.e. passage) context. As we will see in the next chapter, in conversational speech we find the same general nonlinear pattern, but without this dip.

As was indicated above (cf. Figure 4.9), the time of the utterance has a large effect on these articulation rates. The model's estimate, of 0.008 per unit of START time, appears small in Table 4.2 but over the course of the entire reading passage this effect is actually quite large – much larger than the effect of regional differences. This is another peculiar finding, in addition to the dip in the effect of NUMSYLS, from the reading passage articulation rate data. Based on our normal human experiences of speaking and listening, speech does not normally speed up (at least so rapidly) in normal situations. I will return to this when I consider the value of read speech for timing analysis at the end of this chapter.

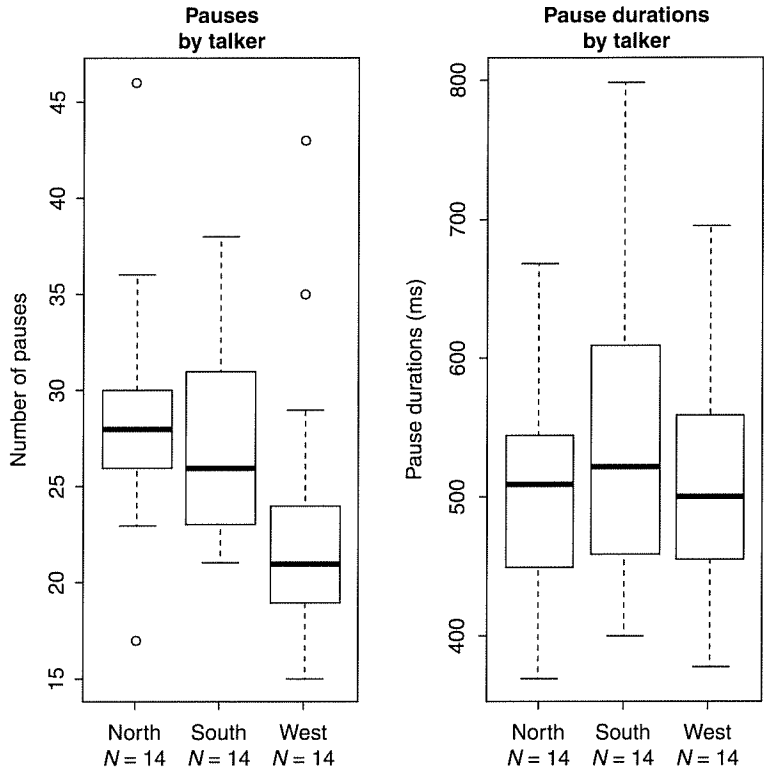


Figure 4.11 Pause *Ns* and pause durations by talker

4.5.2 Pauses in the reading passage data

When we turn our attention to the pauses in the reading passage data, we see a similar regional pattern, where the Western talkers stand out from the Northern and Southern ones. This is clearly visible in Figure 4.11, which shows the number of pauses per speaker in the boxplot on the left. Westerners have a low number of pauses compared to Southerners and Northerners. On the right of the figure, we see the median pause durations across the talkers and observe that the pause durations are roughly similar across the three regions; Westerners are not making up for their fewer pauses by producing longer pauses, they are simply pausing less. Although this pattern is visibly striking in the boxplot for the number of pauses, it does not actually reach significance (ANOVA, $F(2, 39) = 1.97$, $p = 0.15$). In fact, a *t*-test comparing the Westerners' number of pauses to the North and South simultaneously also does not reach

significance ($p = 0.07$). However, coupled with their higher articulation rates, Westerners finish the reading task more quickly than the other regions; the difference between the groups is significant for the total duration of the reading (ANOVA, $F(2, 39) = 4.24$, $p = 0.02$). This likely contributes to the higher significance of the speaking rate ANOVA than the articulation rate ANOVA in the last section.

The pause data, as measured here, are in general less complex than the rate of speech data and there is not much else to say about these pause duration data points. As §4.4.2 and Figure 4.4 indicated, pause durations distribute in a roughly log-normal fashion and, thus, are converted to log-ms for modeling (and, again, Appendix II provides a set of correspondences between ms and log-ms). The statistical analysis for pause duration at the pause level, however, does not yield any significant effects beyond the random intercept for talker and thus I do not present model results, as there are none. As Figure 4.11 indicated at the speaker level, there do not appear to be differences in the pause durations based on region. I have not provided boxplots showing the pause data at the pause level or as organized by the sex of speaker. None of the available comparisons yield significance for pause duration.

4.6 From investigating read data to conversational speech data

The investigation of this chapter was intended to use a small, balanced, and controlled dataset to take a first look at the methods and findings of a speech rate and pause analysis across three regional groups. As countless laboratory studies have established, read speech data provide a nicely controlled setting for investigating various phenomena. However, for investigating aspects of speech timing, like articulation rates and pauses, reading-based tasks may create more confounds than they eliminate.

For instance, we saw here that articulation rates increase over the course of the reading passage and that this was a highly significant factor in the statistical analysis. Yet, based on our normal experiences as speakers and listeners, we would not expect an increase in rate over time to be a part of normal talk interactions and there seems to me no reasonable explanation for it other than as an artifact of the reading task. (None of the conversational data examined in the following chapters shows this effect.) It appears likely that the subjects speed up as they become more familiar with the reading passage and, especially, as they anticipate the end of the passage. As we saw in the "bumpy" curvilinear pattern for the effect of the number of syllables on articulation rate,

there may also be prosodic confounds in the reading passage, causing certain utterances (of a certain length) to be read in a certain way. Again, this seems problematic if our goal is to gain insight into natural talk or the sociolinguistic influences on speech timing features.

Further, despite our best intentions to create a controlled environment by using read, laboratory-based techniques, it is in actuality quite difficult to ensure that reading styles really are equivalent across subjects. This problem is not limited to read or laboratory speech, however. For instance, in their attempt to look at regional speech rates in the US through an examination of classroom presentations and conversational group discussions, Ray and Zahn hoped that the balanced genres they recorded would collect comparable data for analysis. However, they ended their discussion of their study by noting,

> The issue of context is a perplexing one. The contexts in this study were thought to be similar, but there may be some variability which detracts from our desired equivalence. Classroom procedures, instructional norms, and student backgrounds may vary in ways that make context different. While our approach is reasonable given the goals of this study, our findings may be limited to the extent that data was not gathered in identical contexts. (Ray and Zahn 1990: 36)

While Ray and Zahn's issue of context is larger than just read speech versus nonread speech, for the purposes of investigating speech timing, read speech likely cannot solve the problem of comparability and control. When asked to read a passage into a microphone in a lab-based setting, even with instructions to "read naturally," some participants may adopt a pedantic or "reading to children" gait, while others may approach the task as a nuisance to be rushed through as quickly as possible. Further, some people are just more fluent readers than others and reading proficiency is surely a factor in read speech rates and pause patterns. As Jacewicz et al. point out at the end of their recent consideration, "although read text provides a valuable testing ground for examination of speech tempo since all speakers produce the same speech sample, it has serious drawbacks because speakers differ in their reading abilities and their reading styles" (2010: 847).

I argue that conversational sociolinguistic interview speech is preferable over read speech in the investigation of speech timing. We cannot ensure that all interviews are the same, but we can interpret the speech obtained in these interviews as coming from speakers responding to the same sort of conversational task. Speakers may respond to that

task differently, but they do so (at least we hope most of the time) as speakers reacting to a more real-world interactive event. As argued in Shuy, Wolfram, and Riley (1968), sociolinguistic interview *style*, if we care to call it such a thing, is an important speech style for individuals, and the language use in that style is likely reflective of language use in a variety of important situations. Shuy et al. explain their interview data from Detroit as follows:

> It was the feeling of the investigators that the recorded speech was *not quite casual but also not formal.* It was a good example of the speech used by children to adults ... and by adults to respected strangers. It could seldom be considered in-group speech, particularly for teenagers or adults. *It is, nevertheless, one of the most important styles of speech used by Americans, for it is this style in which they make their moves up (or down) the social scale.* (Shuy et al. 1968: 28; emphasis added)

Having set up and described the general procedures for analysis, and motivated the use of sociolinguistic interview speech for a sociolinguistically motivated research project, in the next chapter we turn to examine a much larger dataset drawn from SLAAP's collection of conversational sociolinguistic interviews.

5
Speech Rate and Pause in Conversational Interviews

5.1 Introduction

Having established methods for the analysis of speech rate and silent pause duration, discussed the statistical techniques, and somewhat problematized the use of reading passage speech for determining social variation in speech timing, we turn now to examine a large dataset of recorded talk from conversational, sociolinguistic interviews in order to establish a quantitative picture of speech rate and pause patterns in spontaneous speech. The study presented in this chapter reflects the main analysis of speech rate and pause in this book. The following two chapters will attempt to expand on the insights gained here, with Chapter 8 changing focus from pause and speech rate as the dependent variables of analysis, the objects of inquiry, to tools for the investigation of other variable linguistic phenomena.

While I ended the last chapter by arguing that speech rates and pauses in out-loud reading are less than ideal for developing a characterization of these features in talk, it could still be possible that conversational speech with its lack of controls is too variable to find systematic patterns. Yet, stereotypes, like that of the "slow talking Southerner," must have origins somewhere, and they are unlikely to have originated in read, laboratory speech. Speech rate and pause have often been thought of as components to larger styles of speaking on the part of speakers. So, as mentioned earlier, Tannen views these features as a part of a set of interactional devices (New York Jewish Conversational Style) rather than a part of a specific sociolect (New York Jewish English) – "the most salient features of the style are fast rate of speech, fast turn-taking (i.e. minimal pause between speakers), and loud voices" (Tannen 1985: 103). The question of whether this kind of interactional or stylistic variability